



CONTRACT NO. A833-136  
FINAL REPORT  
MAY 1992

LIBRARY  
CALIFORNIA AIR RESOURCES BOARD  
P.O. BOX 2815  
SACRAMENTO, CA 95812

# Structural Modeling of Epidemiological Time Series

QA  
280  
S48  
1992

**CALIFORNIA ENVIRONMENTAL PROTECTION AGENCY**



**AIR RESOURCES BOARD**  
**Research Division**



# STRUCTURAL MODELING OF EPIDEMIOLOGICAL TIME SERIES

LIBRARY  
CALIFORNIA AIR RESOURCES BOARD  
P.O. BOX 2915  
SACRAMENTO, CA 95812

## Final Report Contract No. A833-136

Prepared for:

Research Division  
California Air Resources Board  
2020 L Street  
Sacramento, CA 95814

Submitted by:

Division of Statistics  
University of California, Davis  
Davis, CA 95616

Prepared by:

R.H. Shumway  
A.S. Azari

**MAY 1992**

### DISCLAIMER

The statements and conclusions in this report are those of the contractor and not necessarily those of the California Air Resources Board. The mention of commercial products, their source or their use in connection with material reported herein is not to be construed as either an actual or implied endorsement of such products.

# STRUCTURAL MODELING OF EPIDEMIOLOGICAL TIME SERIES

## ABSTRACT

We consider the problem of developing a predictive model for mortality in Los Angeles County as a function of temperature and carbon monoxide (CO), a primary pollutant generated by auto emissions. In an earlier study, Shumway et al (1988) analyzed data in both high (less than 10-day periods) and low (more than 10-day) frequency ranges, finding the most significant associations between mortality and temperature and pollution jointly in the low frequency period band. Time series regression models that controlled for autocorrelation and included nonlinear temperature and pollution effects were developed to predict pollution.

The results of the earlier study raised several important questions that were addressed in this current research. The question "Who is dying?" is studied by modeling within different age-groups (0-9 yrs, 45-64 yrs, 65+ yrs) and it is shown that the greatest fluctuations induced by temperature and pollution occur in the oldest age-group. A second question relates to controlling for the different exposure levels in the different areas of Los Angeles County; hence, the present analyses are done separately for the Central-Coastal and Interior-Valley region. The third question answers the speculation that statistically significant effects seen in the lower frequency bands might be due to common seasonal effects or some other common components lurking in temperature and pollution.

In order to answer the above questions, we use a time series structural component model that includes the fixed regressors temperature, squared temperature and pollution along with a random component emulating some possible common disturbed (i.e. noisy) periodicity. It is found that the common unobservable random component is substantial but is not sufficient to explain the observed mortality in any age group. The statistically significant effects of temperature and pollution as measured by carbon monoxide (or any of the other essentially collinear auto-generating emissions) are still present. We compare the results of fitting the above parametric models to the results of simple nonparametric smoothing techniques designed to display mortality in profile as a function of pollution and temperature. The mortality profiles are constructed separately for the Central-Coastal and Interior-Valley sections of LA County for several different age-groups (0-9 yrs, 45-64 yrs, 65+ yrs) and it is shown that there are substantial mortality fluctuations introduced by both temperature and CO gradients, particularly in the over 65 year old group. The shape

of the mortality surface is consistent with a model that predicts mortality to be a function that varies quadratically with temperature and linearly with CO levels.

The consistency exhibited by the parametric and nonparametric profiles means that we can use either to predict the average mortality by age-group and location as a function of temperature and CO level. Using the parametric model, we make detailed predictions of the linear effect of pollution on the mortality rate, expressed in deaths per 100,000 per day per ppm CO. Such predictions were highest for the over-65 age group, ranging from .185 deaths per day in the Central-Coastal section to .100 deaths per day in the Interior-Valley section with standard errors of .018 and .013 respectively. Rates in the 45-64 age group were .021 and .009 respectively for the Central-Coastal and Interior-Valley groups with standard errors of .003 in both. Very small marginally significant effects were found for children under 10 years old amounting to .003 and .004 respectively for the Central-Coastal and Interior-Valley sections, with standard errors of .001 and .002 respectively.

The effects of other pollutants were not evaluated specifically since they were either collinear with carbon monoxide (CO) or temperature. The pollutants measuring particulate concentration (KM) and hydrocarbons (HC) were essentially collinear with CO whereas the effect of ozone (OZ) was essentially collinear with temperature. An earlier study (see Shumway et al, 1988) established that other pollutants such as sulfur dioxide (SO<sub>2</sub>) and nitrogen dioxide (NO<sub>2</sub>) were not statistically significant.

## ACKNOWLEDGEMENTS

We are indebted to Professor Glen Cass of the Environmental Quality Laboratory at California Institute of Technology and to Mr. Marvin Zeldin of Instaweather, Inc. for valuable suggestions made at the onset of the study. We acknowledge a discussion with Peter Wakamatsu of the Los Angeles Department of Health relating to the racial composition of the Los Angeles County study areas. The study would not have been possible were it not for the Herculean efforts of Dr. John Moore of the Research Division, California Resources Board in obtaining the Los Angeles County census tracts and in defining the age by region mortality partition and the regional pollution and weather partitions. Dr. Moore also contributed substantially to the subject matter and form of the final report. Any errors or technical inadequacies in the analysis that remain are our responsibility.

## TABLE OF CONTENTS

1. Background .....	9
2. Preparing the Los Angeles Data .....	12
2.1 Regionally Partitioned Age-Specific Mortality .....	13
2.2 Regional Pollution and Weather Data .....	16
2.3 Data Reduction: Filtering, Subsampling Pollutants and Mortality by Cause ..	17
3. Nonparametric Modeling of Mortality .....	20
3.1 The Winter-Summer Controversy .....	21
3.2 Nonparametric Mortality Prediction by Age and Region .....	22
4. The Possible Role of Time Series Components .....	23
4.1 Mixed Dynamic Linear Models .....	24
4.2 Extraction of Mortality Signals .....	27
4.3 Parametric Mortality Prediction .....	31
5. Summary and Conclusions .....	32
6. References .....	34

## LIST OF FIGURES

	Page
Figure 1: Map of Los Angeles County Partitioned into Central-Coastal and Interior-Valley Sections with Air Quality and Weather Stations.	37
Figure 2: Original Daily Records for Total Mortality, Temperature and Carbon Monoxide (CO) Level (1970-1971).	38
Figure 3: Scatterplots Relating Temperature and CO to Total Mortality (1970-1979, 3652 days).	39
Figure 4: Nonparametric Surface and Contour Plots Relating Total Mortality to Temperature and CO Level (1970-1979, 3652 days).	40
Figure 5: Total Mortality Rate (Deaths/100,000/day) for Central-Coastal and Interior-Valley Sections of Los Angeles Co. (1970-1979, 721 5-day intervals).	41
Figure 6: Temperature and CO Levels for Central-Coastal and Interior-Valley Sections of Los Angeles Co. (1970-1979, 721 5-day intervals).	42
Figure 7a: Filters Applied to Los Angeles Data Are 20 pt. Moving Average (upper panel) and Linear Filter Designed by Shumway et al (1988)(lower panel).	43
Figure 7b: Filters Applied to Los Angeles Data in This Study.	44
Figure 8: Raw, Filtered and Subsampled Filtered Central-Coastal Total Mortality Record (1970-1979, 721 5-day intervals).	45
Figure 9: Raw, Filtered and Subsampled Filtered Central-Coastal Temperature Record (1970-1979, 721 5-day intervals).	46
Figure 10: Scatterplots Relating Respiratory and Cardiovascular Mortality to Total Mortality (1970-1979, 3652 days).	47
Figure 11: Scatterplots Relating CO Used in Study to Other Correlated Pollutants HC and KM for Central-Coastal and Interior Valley Sections (1970-1979, 721 5-day intervals).	48
Figure 12: Scatterplots Relating Ozone and Temperature, for Central-Coastal and Interior-Valley Sections in Los Angeles Co. (1970-1979, 721 5-day intervals).	49
Figure 13: Scatterplots Relating Temperature and Detrended CO to Detrended Mortality for Central-Coastal and Interior-Valley Sections. (1970-1979, 721 5-day intervals).	50

	Page
Figure 14: Comparison of Summer and Winter Mortality Profiles with Temperature and CO (1970-1979, Raw Data, 128 days per Season).	51
Figure 15a: Mortality Rate (deaths/100,000/day) Profiles and Contours With Temperature and CO for Age 65+ for Central-Coastal Section (1970-1979, 721 5-day intervals).	52
Figure 15b: Mortality Rate (deaths/100,000/day) Profiles and Contours With Temperature and CO for Age 45-64 for Central-Coastal Section (1970-1979, 721 5-day intervals).	53
Figure 15c: Mortality Rate (deaths/100,000/day) Profiles and Contours With Temperature and CO for Age 0-9 for Central-Coastal Section (1970-1979, 721 5-day intervals).	54
Figure 15d: Mortality Rate (deaths/100,000/day) Profiles and Contours With Temperature and CO (Total) for Central-Coastal Section (1970-1979, 721 5-day intervals).	55
Figure 16a: Mortality Rate (deaths/100,000/day) Profiles and Contours With Temperature and CO for Age 65+ for Interior-Valley Section (1970-1979, 721 5-day intervals).	56
Figure 16b: Mortality Rate (deaths/100,000/day) Profiles and Contours With Temperature and CO for Age 45-64 for Interior-Valley Section (1970-1979, 721 5-day intervals).	57
Figure 16c: Mortality Rate (deaths/100,000/day) Profiles and Contours With Temperature and CO for Age 0-9 for Interior-Valley Section (1970-1979, 721 5-day intervals).	58
Figure 16d: Mortality Rate (deaths/100,000/day) Profiles and Contours With Temperature and CO (Total) for Interior-Valley Section (1970-1979, 721 5-day intervals).	59
Figure 17: Original Points Superimposed on the 65+ Mortality Profiles for the Central-Coastal and Interior-Valley Populations (Larger dot is data point, smaller dot is label).	60
Figure 18: Original Detrended 65+ Central-Coastal Mortality Compared With Estimated Inherent Mortality Signal Estimated by Model I (middle panel) and Model II (bottom panel) Models.	61
Figure 19: Bootstrap Sampling Distributions of the Coefficients of Temperature ( $\times 10$ ), Squared Temperature ( $\times 100$ ), CO, the Two Estimated Autoregressive Coefficients and the Two Variances.	62

Figure 20a: Total Predicted Mortality (deaths/day) Contours by Age and Section as a Function of Temperature and CO Using Parametric State-Space Model (1970-1979, 721 5-day intervals).	64
Figure 20b: Total Predicted Mortality (deaths/day) Contours by Age and Section as a Function of Temperature and CO Using Parametric State-Space Model (1970-1979, 721 5-day intervals).	65
Figure 20c: Total Predicted Mortality (deaths/day) Contours by Age and Section as a Function of Temperature and CO Using Parametric State-Space Model (1970-1979, 721 5-day intervals).	66
Figure 20d: Total Predicted Mortality (deaths/day) Contours by Age and Section as a Function of Temperature and CO Using Parametric State-Space Model (1970-1979, 721 5-day intervals).	67

## 1. BACKGROUND

Numerous analyses of relationships between daily mortality, pollution and weather in a number of metropolitan areas have been performed to determine if higher levels of pollution are associated with higher mortality. These analyses have employed a wide variety of statistical methods; analyses using time series methods have been the most successful. The London metropolitan area, for example, has been the subject of a number of investigations attempting to link daily observations on mortality with temperature, relative humidity, sulfur dioxide levels and British smoke levels measured over 14 winters. A review of a number of such studies, including the initial study by Martin and Bradley (1960) is given in Ware et al (1981); these early efforts used linear regression to relate mortality to the primary pollutants for the winter of 1958. Later studies such as Mazumdar et al (1982), Ostro (1984), Shumway (1983), Schwartz and Marcus (1990) have used various linear or nonlinear regression techniques on the entire 14 winters. These studies involving the London data find significant associations between mortality and British smoke and between mortality and sulfur dioxide in the presence of temperature effects.

Another population of great interest that has been persistently exposed to ambient pollution concentrations exceeding health-based standards is that of the greater Los Angeles area. Because the pollutants of primary importance are those commonly associated with automobile emissions, the effects of pollutants such as carbon monoxide (CO), hydrocarbons (HC) and particulate concentrations (KM) become important. Early pioneering studies such as Mills (1960) and Hechter and Goldsmith (1961) reported significant (or nonsignificant) associations of mortality with ozone and CO levels. The years 1962-1965 were analyzed using multiple regression methods by Hexter and Goldsmith (1971) who "identified" a final model involving temperature and log CO. Seasonality was modeled in terms of trigonometric component terms in the regression.

Two recent studies reported by Shumway et al (1988) and Kinney and Özkaynak (1991) have focused on analyzing the daily deaths in Los Angeles County for the years 1970-1979. Using time series nonlinear regression techniques, Shumway et al (1988) searched for linear effects on total, cardiovascular and respiratory mortality that might involve combinations of temperature, relative humidity and six primary pollutants, Carbon Monoxide (CO), Sulfur Dioxide (SO<sub>2</sub>), Nitrogen Dioxide (NO<sub>2</sub>), Hydrocarbons (HC), Ozone (O<sub>3</sub>) and particulates (KM). They performed extensive stepwise regressions and reported results for 11 different frequency bands encompassing both short and long periods.

Shumway et al (1988) also found some statistically significant effects at short periods

of 2-30 days that were of little practical importance. Kinney and Özkaynak (1991) performed regression analyses on the high frequencies which they extracted using the broadband high-pass filter developed in Shumway et al (1988) to control for seasonality. They reported statistically significant coefficients using a model involving (1) lagged O<sub>3</sub>, NO<sub>2</sub> and temperature, (2) lagged O<sub>3</sub>, CO and temperature and (3) lagged O<sub>3</sub>, KM and temperature. Again, these results, while statistically significant, never accounted for more than 4% of the total variation, implying that they are practically insignificant. Shumway et al (1988) found in this frequency range that while temperature alone was statistically significant, including pollution in the model contributed little to explanatory power.

In addition to the high frequency analysis, Shumway et al (1988) included an analysis of the lower frequency effects that included those frequencies corresponding to seasonal oscillations in temperature. They found that a linear model involving temperature and one of the auto-generated pollutants (CO, HC or KM) accounted for substantial portions of variation (69-82%) in a broad band centered on 32 weeks. They went on to construct a file of filtered data restricted to those lower frequencies and found that there were significant nonlinearities present that tended to be quadratic in temperature and logarithmic in pollution. This implies that substantially more people die on both hotter and colder days. They obtained nonparametric dose response profiles that predict the average number of additional deaths due to changes in temperature or in the level of one of the primary pollutants. They also fitted a parametric nonlinear model that controlled for autocorrelation through an autoregressive error term. Both models led to comparable gradients in mortality as a function of temperature and a primary pollutant. Typical percentages of variation accounted for by ordinary linear regressions were in the neighborhood of 35-40%.

Kinney and Özkaynak (1991) point out that any effects that might be found within these frequencies are most likely not of biological origin. In their opinion, it is not plausible from a biological point of view that more die in the winter because of the colder temperatures prevalent during that time of the year. On the other hand, the effects of excessive heat are presumably accepted after the study of Oechsli and Buechley (1970). Since our earlier (1988) analysis found excessive deaths both during cold periods and during hot periods in the seasonal and neighboring frequency bands, we felt that additional efforts focusing on these lower bands would be desirable. We show in this study that there is a substantial component in the mortality series that rises and falls independently of temperature with the seasons. We will also show that there are statistically significant effects of temperature on mortality that are quite separate from this underlying unobservable seasonal mortality signal.

One of the primary objectives of this project was to attempt to isolate the time series component mentioned above that could not be explained by temperature and pollution fluctuations. We will present first a simple analysis that involves separating out winter and summer files; we took 128 days in each of the two seasons for each of the years 1970-1979 and analyzed the files separately. The second approach to the seasonality problem involves comparing parametric models that contain an unobserved seasonal component to models that add the covariates temperature and CO level to this unobserved component. Using this simple structural model, the unobservable mortality signal can then be separated from the other effects and assessed as a seasonal waveform. We also compare the parametric predicted means with those that are obtained via nonparametric smoothing.

The second objective of this study is to investigate the mortality profiles as a function of age and of location of primary residence. The county of Los Angeles is divided into a Central-Coastal population and an Interior-Valley population using 1980 census records. This means that elements of the two populations have a better chance of being associated with the pollution and temperature levels that they actually experienced. The population is partitioned further by age and we present the analysis separately for the groups of interest (0-9, 45-64, 65+) to determine whether there are differential mortality gradients induced as a function of age.

## 2. PREPARING THE LOS ANGELES DATA

We make use of two different data files in this study for the purpose of assessing different models and techniques. The *Initial Study Raw Data File* involved (see Shumway et al, 1988) 11 series of daily measurements in Los Angeles County spanning a ten-year period (1970-1979) and consisting of 3652 daily observations on three mortality series (Total Mortality, Respiratory Mortality and Cardiovascular Mortality) where Total Mortality excludes deaths due to accidents and suicides. The raw meteorological data in that study consisted of maximum daily temperature and average relative humidity averaged over Downtown Los Angeles and the Los Angeles International, Long Beach, Burbank and Ontario airports. The pollution data consisted of the average of the daily one-hour maxima of six pollutants, carbon monoxide (CO), sulfur dioxide (SO<sub>2</sub>), nitrogen dioxide (NO<sub>2</sub>), ozone (O<sub>3</sub>), total hydrocarbons (HC) and the (KM) measure of particulate concentrations at six monitoring stations. The monitoring stations were located in Azusa, Burbank, Downtown Los Angeles, North Long Beach, Reseda and West Los Angeles.

It is useful to mention in passing that the coefficient of mass (KM) is a measure of soiling capability of the airborne particulate matter, calculated as a simple logarithmic function of the percent transmission from the soiled spot with respect to the transmission from the clean paper. In the KM monitoring stations, ambient air was pulled through a segment of porous tape during 2-hr intervals and then the amount of light transmitted through the tape was recorded to evaluate the KM measure. The method is quite sensitive to particles small enough to penetrate deep into human lungs and produce possible adverse health effects. Detailed characteristics of Los Angeles particulate matter in this size range and its composition are given in Conklin, Cass and Macias (1982), Cass and Conklin (1984), and Gray and Cass (1986).

For the purpose of this study we partitioned the *Initial Study Raw Data File* by region and age. This yields 3652 days of mortality, pollution and weather values classified by region (Central-Coastal or Interior-Valley) and by age (0-9, 45-64, 65+). The two regional files each contained 12 series with Total Mortality for all ages and for the groups 65+, 45-64, 0-9 comprising the first four series and temperature, relative humidity and the six pollutants making up the remainder.

In Shumway et al (1988), the analysis showed that most of the predictive associations between the weather, pollution and mortality series tended to be in a low frequency signal band composed of periods greater than 10 days. The frequencies corresponding to higher frequencies than these essentially represent noise since pollution levels and temperature

had low correlations with mortality at those frequencies. The higher frequencies were filtered out using a low pass filter and the output was subsampled at 5-day intervals with essentially no loss of information. This yields a *Filtered Central-Coastal File* and a *Filtered Interior-Valley File* containing the same 12 series in the order mentioned above.

Section 2 gives our motivation for considering the above partitioned and filtered files and details of their construction. The reader willing to accept these files at face value can proceed directly to Section 3.

## 2.1 Regionally Partitioned Age-Specific Mortality

The regional and age-specific mortality series were extracted from an extensive mortality file including all deaths of Los Angeles residents and nonresidents in Los Angeles County during the 10 year-period 1970-1979. The file also includes deaths of Los Angeles residents which occurred in some other locality. The International Classification of Disease (ICD) Codes, eighth and ninth revisions, were used to classify total mortality into respiratory and cardiovascular mortality. The ICD codes were also used to eliminate the cases with the cause of death being classified as accident, suicide and homicide.

To eliminate the effect of averaging a mortality response function over distinctly different conditions, Los Angeles County was subdivided into Central-Coastal and Interior-Valley regions. Deaths of residents of desert regions of the county north of mountains surrounding the Los Angeles Basin were omitted from the database. Since the census tracts of the last residences of the subjects were to be used to assign subjects to regions, the boundary separating the regions was drawn along the boundaries of census tracts. Because the database contained assignments to both 1960 and 1970 tracts, slightly different boundaries were drawn for the two kinds of tracts. The boundaries ran eastward along the crest of the Santa Monica Mountains and thence southeast along the western bases of the Monterey Hills and Puente Hills to the southern boundary of the county. A boundary in this vicinity was suggested by Glenn Cass and Marvin Zeldin as a good approximate separation between the Central-Coastal region with lower ozone concentrations and the Interior-Valley region with higher ozone concentrations.

The Central-Coastal region includes downtown Los Angeles; the Interior-Valley region includes the San Fernando and San Gabriel Valleys. The boundary approximately coincides with a portion of the of a 20-day isopleth on the map "Number of Days Exceeding Stage One Episode Level (0.20 ppm)" in the summary of 1979 air quality in the South Coast Air Basin (Shikiya et al, 1980). The boundary and the weather and pollution monitoring

stations used in the study are shown in Figure 1.

To study the effect of the air pollution on different age groups, age-specific mortality files were created for each region using the age codes of the mortality file. The age groups of interest were 0-9, 45-64 and 65 and over. Table 1 shows the population by age for the Central-Coastal and Interior-Valley regions defined by the partition shown on Figure 1. The census values for 1980 were used; these are obviously not the ideal figures for converting the total deaths per year to a rate per 100,000, but one cannot recover the year-by-year values and this ending value will give an approximation. The change in population over the interval can be subsumed into the time detrending term in our regressions and it is hoped that gross changes in age or regional composition over the time period were minor.

About 60% of the population resided in the Central-Coastal Region and about 40% resided in the Interior-Valley Region. About 10% of the total are in the over 65 age group where changes in temperature and levels of air pollution induced the largest changes in daily mortality. Members of the population who were 45-64 or 65 or over comprised about 30% of the total in both regions. These last two groups also appeared to have the highest mortality gradients associated with changes in pollution and temperature. The lowest age category studied (0-9) made up only about 15% of the total and had very small insignificant increases in mortality associated with increases in pollution.

Table 2 shows the detrended (over time) mean mortality rates in deaths per 100,000 per day for the *Filtered Central-Coastal File* and *Filtered Interior-Valley File*. In this study, mortality rates are expressed in deaths per 100,000 per day. The mean mortality rates of the filtered series for the two regions are also displayed in Table 2. Note the generally higher mortality rates in the Central-Coastal Region. These differentials are likely to be primarily due to the different racial compositions in the two regions. For example, the overall age adjusted mortality rates for Hispanics, Whites and Blacks in Los Angeles County were 1.37, 2.13 and 3.29 deaths per 100,000 per day respectively (Wakamatsu, 1991).

**Table 1: Subdivision of the Population of LA County**  
1980 Census Tracts\*

Age Group	Central-Coastal	Interior-Valley	Total
0-9	643,051	441,015	1,084,406
10-44	2,438,449	1,754,440	4,192,889
45-54	427,368	333,448	760,816
55-64	396,710	299,913	696,623
65-74	260,819	185,436	446,255
75-84	172,402	124,329	296,730
<b>Total</b>	<b>4,338,854</b>	<b>3,138,649</b>	<b>7,477,503</b>

\* Obtained from Demographic Research Unit, Dept. of Finance, State of California, 915 L St., Sacramento, CA 95814.

**Table 2: Means (Time Trend Adjusted)**

Category	Central-Coastal	Interior-Valley
Total Mortality*	2.10	1.44
65+ Mortality	14.73	10.36
45-64 Mortality	2.91	1.81
0-10 Mortality	.56	.37
Weather Effects:		
Temperature	72.07	77.18
Relative Humidity	62.36	30.93
Pollution Levels:		
CO	8.28	7.63
SO2	3.47	2.21
NO2	11.71	10.71
Hydrocarbons	44.78	54.43
Ozone	6.06	10.45
KM	48.51	46.40

\* Rate per 100,000 per day

## 2.2 Regional Pollution and Weather Data

The two weather series, maximum daily temperature and average daily relative humidity, were extracted from a file containing data for Downtown Los Angeles and the four airports, Los Angeles International, Long Beach, Burbank and Ontario. The averages of the data from Downtown Los Angeles, Los Angeles International Airport and Long Beach were used as the weather series for the Central-Coastal region. The averages of the two remaining stations at Burbank and Ontario were used for the Interior-Valley region.

The six pollutants, namely Carbon Monoxide (CO), Sulfur Dioxide (SO<sub>2</sub>), Nitrogen Dioxide (NO<sub>2</sub>), Total Hydrocarbons (HC), Ozone (O<sub>3</sub>) and Particulate (KM) were measured daily at six stations located in Azusa, Burbank, Downtown Los Angeles, North Long Beach, Reseda and West Los Angeles. The values used for the Central-Coastal region are the average of the daily maxima for the three stations at Downtown Los Angeles, North Long Beach and West Los Angeles. The pollution data for the Interior-Valley region are the average of the daily maxima for the remaining three stations at Azusa, Burbank and Reseda.

Note in Table 2 the generally higher temperature mean in the Interior-Valley and the higher levels of HC and O<sub>3</sub>; the mean mortality rates for the Interior-Valley Region are somewhat lower. The higher mean levels of CO in the Central-Coastal are suggestive since the over all levels of mortality are greater. These data will be discussed further in later sections.

## 2.3 Data Reduction: Filtering, Subsampling Pollutants and Mortality by Cause

In this section, we give details motivating the reduction of the overall data to the filtered files that we present in our analyses. Figure 2 shows a portion (1971-1972) of the daily records corresponding to total daily mortality, temperature and CO; these are the original data records (*Initial Study Raw Data File*) appearing in our earlier study (Shumway, et al, 1989). The yearly seasonality is clear for these three series; the question is whether the fluctuations in total mortality are more closely associated only with the common driving seasonal pattern or whether they are also associated with fluctuations in temperature and pollution after the common seasonal pattern has been extracted to the extent possible.

One can begin to isolate some plausible relationships by looking at scatterplots relating the temperature and CO levels to total mortality. Such scatterplots, shown in Figure 3, lend credence to an association between temperature and mortality that may bend up at both the higher and lower temperatures. The association with CO level (if it exists) tends to be more linear or even to be linear on a logarithmic scale. Of course, the scatterplots are very rough and one would like to do some smoothing over the temperature-pollution plane to isolate plausible functional relations. The relatively low correlation between temperature and mortality is due to the nonlinear effect that can be noticed on the scatterplot.

Figure 4 shows a three-dimensional plot of mortality as a function of temperature and CO level. The plotted mortality values are heavily smoothed versions of daily mortality values. The smoothed values were computed at the points of a uniform 40x24 grid of temperature x CO level values by a two-step process. The first step computed the weighted average of a set of mortalities in the neighborhood of a gridpoint with weights proportional to inverse squared distance. Neighboring points were added to the set in order of increasing distance from the gridpoint until either the set contained ten points or the Euclidean distance between the gridpoint and the next closest point exceeded 50. The second step computed the unweighted average of these smoothed values over the rectangular (9x5) array of neighboring gridpoints.

Two depictions of the surface are useful for representing the mortality profiles as functions of temperature and CO level. The first, shown in the top frame, is essentially a surface plot and displays the general functional form of the surface as being quadratic in temperature and linear or possibly nonlinear in CO, with the CO nonlinearity tending to look fairly logarithmic. These observations led to fitting a nonlinear parametric model in Shumway et al (1989) that was quadratic in temperature and logarithmic in the pollution

direction. The high mortalities at high temperatures and CO levels are not fitted well by the above simple model; adding interactions did not improve the fit. Since such departures from an expected smooth surface may be due to end effects because of averaging over a very few points, we will also fit parametric models to these data in a later section.

The topographic rendition in the bottom frame enables us to make rough numerical assessments relating the changes in mortality to associated changes in temperature and CO level. For example, at a temperature of 80 degrees, a doubling from the CO level of 10ppm to 20ppm would be predicted to increase the number of deaths from approximately 2.28 per day to 2.48 per day or a predicted gradient of .02 deaths per 100,000 per ppm CO per day. These profiles confirm the value of averaging daily mortality over adjacent temperature and CO levels to estimate a mean response function and we adopt these nonparametric profiles as a method for displaying the relationships within these populations.

The noisy character of the raw series shown in Figure 2 suggests the possibility of smoothing over time to eliminate oscillations not likely to be shared by the mortality, weather and pollution series. In addition, filtering and subsampling will reduce the volume of data to be considered by about 80%. Filtered renditions of the daily mortality values are shown for comparison in Figure 5. It was established in the earlier study that this kind of smoothing does not eliminate any fluctuations in mortality that are associated with changes in temperature or CO levels. Figure 6 shows comparable levels of temperature and CO level in the two regions. The data are reduced in this way from 3652 daily values to 721 values taken at 5-day intervals without changing the character of the series.

The filters chosen were a refinement of those used in the earlier study. In general, earlier analyses showed that the frequency dependent correlation or coherence was highest in frequency bands corresponding to periods greater than 12-15 days; two filters that tend to eliminate the noise in higher frequencies not including these periods are shown in Figure 7a. In this study, we chose to let in some more of the higher frequencies (lower periods) by re-designing a filter with the frequency response shown in Figure 7b. This led to the *Filtered Central-Coastal File* and the *Filtered Interior-Valley File*. Results of applying these filters are shown in Figures 8 and 9. Filtering eliminates the higher frequency noise and the subsampling reduces the volume of data that needs to be considered. Contours generated for the filtered file hardly differ from those given using the original raw data as shown in Figure 4 (compare Figure 15d).

The use of total mortality as a surrogate for either cardiovascular or respiratory mortality needs some comment. The relevant scatterplots along with correlations are shown

in Figure 10, where we see that the correlation between total and cardiovascular mortality is 0.83 and that the correlation between total mortality and respiratory mortality is lower (0.50), due to the small number of cases of respiratory mortality. The correlation with respiratory mortality is also essentially linear with total mortality resulting in a correlation of 0.71. Cross correlation analysis of these mortalities shows no leading or lagging effects. To eliminate duplication of tedious computer runs, we restrict our analysis to the total mortality series.

We shall also use the CO level as a surrogate for other pollutants due to primary auto emissions (HC and KM) rather than repeating runs for each of them separately. Scatterplots of HC and KM against CO in Figure 11 show high correlations (.80-.86) and we infer that they are essentially collinear. While others have advocated (see Kinney and Özkaynak, 1991) including temperature and ozone levels (OZ) in the model at the same time, our previous study indicated that OZ comes into the model after temperature has already been added and that OZ is essentially collinear with temperature. As a rough justification for this assertion, we offer the scatterplots in Figure 12 which show the relation between temperature and OZ in the Central-Coastal and Interior-Valley files. Note that the correlations are .78 and .89 respectively. For a more careful analysis, see Shumway et al (1988).

The above comments are made to motivate the reduction of our data to the subseries, total mortality, CO level and Temperature and to justify the use of the filtered and sub-sampled series. This brings the problem of analyzing the data by age and region down to a manageable level without ignoring potential contributing variables.

As a final remark, note that there are often trends over time in the mortality and CO series. In Figure 5, the Central-Coastal mortality shows a decreasing linear trend. Similar decreasing trends in CO levels are shown in both the Central-Coastal and Interior-Valley files plotted in Figure 6. In order to discount the possibly spurious effects that may be due to other factors changing over time, we will analyze detrended series. Scatterplots involving detrended series, shown in Figure 13, are quite similar to the raw data scatterplots shown in Figure 3.

### 3. NONPARAMETRIC MODELING OF MORTALITY

The nonparametric mortality profile as shown in Figure 4 is an appealing way of summarizing the information relating temperature and any pollutant to daily mortality. The regularity of the surface gives an intuitive feel for the nonlinear variation of mortality as a function of the inputs and the topographic rendition in the lower half allows one to read quantitative values of mortality for each combination of a temperature and pollution value. There are numerous choices for the methods that might be used in the smoothing procedure. In this report, we have experimented with various combinations of nearest neighbor and rectangular averaging (kernel) smoothing as discussed at the beginning of Section 2.3.

The smoothed contours described above involve averaging values of mortality in temperature-pollution neighborhoods without regard for their time positions in the record. It is not completely clear that this eliminates the contaminating effect of time trends or seasonality from consideration. It would be tedious to identify uniquely the position in time (winter or summer, for example) where each point in any nearest neighbor average came from. In Section 3.1 below, we attempt to provide a partial answer to this question by constructing separate winter and summer plots and comparing them with a plot of all data. A more complete answer is given in Section 4 where we find that a time series components model yields the same predicted mortalities as the nonparametric ones derived in Section 3.2 below.

In the meantime, we note that the scatterplots relating the inputs temperature and CO (after time detrending) tend to confirm the basic pattern noticed in the earlier data. Figure 13 shows scatterplots for both the Central-Coastal and the Interior-Valley filtered files and we see evidence of the same quadratic pattern with temperature and the linear or logarithmic pattern with CO level.

### 3.1 Resolving the Winter-Summer Controversy

An approach to resolving the purely seasonal effects on mortality is to construct separate winter and summer files as described at the beginning of Section 2. Time plots of these separate seasons (not shown in this report) are relatively flat over the winter (November 1-March 8(7)) and over the summer (June 1-Sept. 30) since the intervening spring and fall values have been deleted. In terms of coverage of the calendar year, the winter file is then roughly comparable to the 14 winters available for the London data (see Shumway et al, 1983).

We show in Figure 14 the result of constructing the nonparametric profiles with temperature and CO level separately for winter and summer and the result of constructing an overall nonparametric profile for the combined data. Essentially the same overall profile results are obtained for the entire raw data file (compare with Figure 4). It is also fairly clear that hot days during the winter when pollution is high also lead to increased mortality. Since there were no really cold days during the summer, we are not able to see the opposite effect during that season.

We conclude on the basis of this rather crude analysis that there is not yet any indication that the mortality gradients are not due to temperature effects but are induced by some as yet unidentified seasonally varying sources. Section 4 examines this latter possibility in greater detail using a parametric structural model with time series components.

### 3.2 Nonparametric Mortality Prediction by Age and Region

We summarize in this section the results of applying the nonparametric profile modeling to mortality in the Central-Coastal and Interior-Valley regions as a function of temperature and CO. Total mortality is expressed as deaths per 100,000 per day. Surfaces and contour plots are displayed for all age groups combined and for the age-groups 65 and older (65+), 45-64 and 0-9.

Results from the Central-Coastal region are summarized in Figures 15(a-d). CO levels are in parts per million (ppm). Examining Figure 15 shows the same sort of quadratic contours in temperature and linear or loglinear relation with CO levels that we have observed before. Notice that the topographical rendition in the lower frame enables reading the mean gradients in mortality predicted for different temperature and CO level gradients. This can be done roughly by reading the mortality differential between the CO reference points 10 and 20 ppm and then dividing by 10 to express the increase as deaths per 100,000 per ppm per day.

If we make the above approximations for the Central-Coastal group, we obtain predicted increases of .18, .03 and .01 deaths per 100,000 per ppm per day in the 65+, 45-64 and 0-9 age groups respectively. For the Interior-Valley region, shown in Figures 16(a-d), the comparable estimates are .08, .01 and .01 deaths per 100,000 per ppm per day in the 65+, 45-64 and 0-9 age groups.

One should use extreme caution in making too many definitive statements on the basis of these nonparametric contours. No confidence intervals are provided for these nonparametric contours; confidence intervals will be given for the parametric model considered later. Some of the regions of the temperature-CO plane had very few points, particularly at high temperatures and pollution levels and at low temperatures and high pollution levels. Figure 17 shows the contours in the two groups with the points superimposed; it is clear that certain regions will generate profiles based on data values that are quite far away. Note the large number of summer observations in the Interior-Valley region with high temperatures and low CO pollution levels.

#### 4. THE POSSIBLE ROLE OF TIME SERIES COMPONENTS

The fact that the observed mortality series display some of the same kinds of periodicities as the temperature and CO series is a possible indicator that the relations between mortality, pollution and temperature we are observing on the nonparametric profiles are contaminated by some unassigned seasonal causes which cannot be specifically observed. Whatever the effects on mortality of these unassignable causes may be, they can only be observed together with the effects of pollution and temperature. These seasonal fluctuations in mortality should be extracted from the data to ensure that our predictions are not distorted by them.

There are a number of simple methods for isolating a possible periodic common signal from the mortality, temperature and pollution series. For example, one might assume that the underlying signal is a simple linear combination of sine and cosine functions with a fixed underlying frequency  $\nu$ . That is, denoting detrended mortality at time  $t$  by  $M_t$ , write

$$M_t = c_1 \cos(2\pi\nu t) + c_2 \sin(2\pi\nu t) + e_t$$

where  $c_1$  and  $c_2$  are the unknown constants multiplying the periodic part that models the common seasonality. Early investigators such as Hexter and Goldsmith (1971) used this technique to control for seasonality.

More recent research has indicated that simple sines and cosines are not sufficient to account for this kind of behavior in real data; Newton et al (1991) show an example involving the fitting of global ice volume measurements where the residuals from models involving sines and cosines still have periodic components near the seasonal frequencies due to the fact that real periodic oscillations are disturbed by addition of a linear combination of shocks where the coefficients are exponentially weighted sines and cosines at a given frequency. They show that autoregressive models are ideal for evaluating disturbed periodicities and essentially model each single periodicity in terms of a second order autoregressive process. Hence, another approach to the model defined above would be to substitute an unobserved second-order autoregressive component for the fixed linear combination of sines and cosines.

We would then obtain

$$M_t = x_t + v_t \tag{Model I}$$

for the detrended mortality where

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + w_t$$

is the unobserved autoregressive component satisfying the usual difference equations. Here  $v_t$  and  $w_t$  are two independent errors associated with this model. We then have a model that says that the detrended mortality series is simply an unobservable signal and may not be influenced by temperature and CO levels. We shall refer to this null model as *Model I*. The unknown parameters in this model are  $\phi_1$ ,  $\phi_2$  and the variances of  $v_t$  and  $w_t$ , say  $\sigma_v^2$  and  $\sigma_w^2$ . The material in Section 4.1 discusses the technical details of obtaining estimators for the unknown parameters and for estimating the unobserved mortality signal  $x_t$ . We only give the results of that procedure here.

The second alternative description that we will look at adds to *Model I* the effects of temperature  $T_t$  (linear and quadratic) and pollution  $P_t$  to obtain the proposed model

$$M_t = c_1 T_t + c_2 T_t^2 + d P_t + x_t + v_t \quad (\text{Model II})$$

where  $x_t$  is the autoregressive component given above and the unknown parameters are  $c_1, c_2, d, \phi_1, \phi_2, \sigma_v^2$  and  $\sigma_w^2$ . The choice of a term that is linear in CO level  $P_t$  may seem somewhat arbitrary because of the plausibility of an alternate explanation in terms of  $\log P_t$ . We used the linear form primarily for ease in exposition so that the coefficient  $d$  would have a simple interpretation in terms of quoted pollution levels. Additionally, there seemed to be little to choose between the two models from the nonparametric contours estimated earlier. The main difference between the two models occurs at high levels of pollutants which are relatively unlikely to occur in the future. We can compare the goodness of fit of *Model II* to that of *Model I* to decide whether temperature and pollution contribute significantly to mortality.

To summarize, we will compare the fit of *Model II* to that of *Model I* to test whether a model with temperature and pollution effects does better than a model which ignores those effects. The unobserved mortality signal  $x_t$  will be estimated under both models to compare its contributions with and without pollution and temperature effects. Finally, we will estimate the predicted mean mortality under the best model and compare this predicted mortality profile to the profiles that we have generated by the nonparametric nearest neighbor smoothing procedure. This analysis will be performed for each of the age groups and total in both the Central-Coastal and Interior-Valley regions.

## 4.1 Mixed Dynamic Linear Models

We summarize briefly the statistical considerations that are involved in maximum likelihood estimation and estimation for *Model II*. The reader willing to accept the estimated parameters and standard errors in Table 3 at face value may proceed directly to Section 4.2. Our notation for the state space formulation of the mixed dynamic linear model follows that in Shumway (1988, Sections 3.4.1-3.4.4) and we will refer to equations in that treatment.

First, note that a general form including *Model II* as a special case is the mixed dynamic linear model

$$\mathbf{y}_t = B\mathbf{z}_t + A\mathbf{x}_t + \mathbf{v}_t,$$

where  $\mathbf{y}_t$  is the  $qx1$  vector of observed values,  $\mathbf{z}_t$  is an  $rx1$  vector of observed covariates and  $\mathbf{x}_t$  is a  $px1$  vector of unobserved random signals. The model is mixed because it consists of fixed and random components and it is dynamic since it evolves in time. The dependence of the observed values on the covariates is modeled by the  $qxr$  matrix  $B$ , and their dependence on the unobserved random signals is modeled by the  $qxp$  matrix  $A$ . The added term  $\mathbf{v}_t$  is the  $qx1$  vector of observation noise, assumed to be independent and normally distributed with mean zero and covariance matrix  $R$ .

In our *Model II*,  $\mathbf{y}_t = M_t$ ,  $\mathbf{z}_t = (T_t, T_t^2, P_t)'$  and  $\mathbf{x}_t = (x_t, x_{t-1})'$ , the vector composed of the signal and its lagged one version. The matrices  $B = (c_1, c_2, d)$  and  $A = (1, 0)$  complete the description of the model for the observed mortality. The unobservable mortality signal is assumed to satisfy the vector state equation

$$\mathbf{x}_t = \Phi\mathbf{x}_{t-1} + \mathbf{w}_t$$

where

$$\Phi = \begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix}.$$

The  $\mathbf{w}_t$  are assumed to be independent and normally distributed with covariance matrix  $Q$ . In our case the covariance matrices are  $R = \sigma_v^2$  and

$$Q = \begin{pmatrix} \sigma_w^2 & 0 \\ 0 & 0 \end{pmatrix}$$

with  $\sigma_v^2$  and  $\sigma_w^2$  denoting the observation and model variances respectively.

The equations above are the same as in Shumway (1988) with the exception of the term  $Bz_t$  involving the covariates. We note that we are interested in estimating the parameters in the model, which are the temperature coefficients  $c_1, c_2$ , the pollution coefficient  $d$ , and noise variances  $\sigma_v^2$  and  $\sigma_w^2$ . We are also interested in estimating the form of the unobservable mortality signal  $x_t$  using the data.

The second problem mentioned above is that of smoothing the observed mortality; this problem can be solved using a modification of the Kalman filter and smoother equations given in Shumway (1988). In general, we define the estimated signal  $\mathbf{x}_t$  based on the data  $y_1, \dots, y_s$  as

$$\mathbf{x}_t^s = E\{\mathbf{x}_t | y_1, \dots, y_s\}$$

with the covariance function given by  $P_t^s$ . We will be primarily using expressions of the type  $s = t - 1$ , defined as filtering or forecasting and  $s = t$  defined as smoothing or interpolation. The estimators for the mortality signal given in the next section are the smoothed estimators. We note here that filtering and smoothing as in Shumway (1988, Section 3.4.2) proceeds in the case of the mixed model by replacing the residuals  $\epsilon_t$  by

$$\epsilon_t = y_t - Bz_t - Ax_t^{t-1}.$$

Maximum likelihood estimation proceeds as in Section 3.4.3 with the replacement of  $y_t - Ax_t^n$  by the new residual  $y_t - Bz_t - Ax_t^n$  and uses the additional equation

$$\hat{B} = \sum_{t=1}^n (y_t - Ax_t^n) z_t' \left( \sum_{t=1}^n z_t z_t' \right)^{-1}$$

for updating the fixed regression coefficients by the EM algorithm.

To develop standard errors for the estimated parameters, we use the bootstrap procedure given by Stoffer and Wall (1991). If we define the generic vector of maximum likelihood estimators obtained above as  $\hat{\Theta}$ , we simply resample the scaled residuals  $\Sigma_t(\hat{\theta})^{-\frac{1}{2}} \epsilon_t(\hat{\theta})$  and reconstruct the data using the Kalman filter equations given in Shumway (1988) evaluated at the rescaled residuals  $\Sigma_t(\hat{\theta})^{-\frac{1}{2}} \epsilon_t^*(\hat{\theta})$  where the bootstrap resamples are denoted by  $\epsilon_t^*$ . Each of  $m$  bootstrap resamplings gives a new set of maximum likelihood estimators, denoted by  $\hat{\Theta}_1^*, \dots, \hat{\Theta}_m^*$  where  $m$ . The sampling distribution of the  $\hat{\Theta}_i^*$  approximates the sampling distribution of the maximum likelihood estimators. The standard deviation

of the bootstrap replicates estimates the standard deviation of the maximum likelihood estimators. These estimators of the standard deviations, based on  $m = 200$  bootstrap replications, are shown in Table 3 of the next section.

A second question of interest is selecting the more correct model. This reduces in our case to choosing between *Model I* and *Model II*. For this test we need the maximized log likelihood, say  $\log L(\hat{\Theta})$ , under each of the models. A straight forward comparison of the models can be made by the likelihood ratio test, whose test statistic is the difference of the values of  $-2 \log L(\hat{\Theta})$  for the two models. This difference is then compared to a chi-squared random variable with degrees of freedom equal to twice the difference between the numbers of parameters in the two models. We prefer to use the classical model selection criterion of Akaike, defined as

$$AIC = -2 \log L(\hat{\Theta}) + 2 \times (\text{number of parameters}),$$

and choosing the model with the minimum *AIC* value (see, for example, Shumway, 1988).

## 4.2 Extraction of Mortality Signals

One of the objectives of this study was to investigate the possibility that the isolated mortality signal description shown as *Model I* is adequate for describing the Los Angeles mortality series. This would mean that adding the covariates temperature and(or) CO pollution would not improve the fit of the model and would imply that these factors were not significant contributors to mortality. In order to test formally whether *Model I* or *Model II* provides the better representation for the mortality data, we can either compare the *AIC* values of the two representations, or we can examine the parameter estimates and their standard errors to see whether we reject the hypothesis that each given parameter is zero.

Consider applying the first procedure of the two mentioned above to the Central-Coastal data for the over 65 age group. The estimated parameters under the *Model I*, computed by the method described in Section 4.1 are  $\hat{\phi}_1 = .90$ ,  $\hat{\phi}_2 = -.11$ ,  $\hat{\sigma}_v^2 = .74$ ,  $\hat{\sigma}_w^2 = 0$ , and the value of the log likelihood is -253. The log likelihood computed under *Model II* was -218. The values of the *AIC* criterion defined in Section 4.1 are  $-2(-253) + 2(4) = 514$  and  $-2(-218) + 2(7) = 450$  for *Model I* and *Model II*, respectively, implying that *Model II* is superior. It is interesting to compare the estimated mortality signals of both models, as in Figure 18. We see that the estimated signal looks very similar to the total mortality series shown in Figure 8 with the exception that the time trend is absent (as it

should be) and that it is smoother. The amplitudes of the mortality signal estimated by *Model II* are about 1 death per 100,000 per day smaller. This decrease in amplitude is approximately the portion of mortality explained by the temperature and pollution effects.

**Table 3: Estimated Parameters (Std. Err.) of Model II**

Age	65+	45-64	0-9	Total
<b>Central Coastal</b>				
<i>Constant</i>	29.166	5.780	.647	3.688
<i>Temp</i>	-.418(.067)	-.069(.020)		-.044(008)
<i>Temp</i> <sup>2</sup> × 10 <sup>4</sup>	26.20(4.5)	4.271(1.4)		2.989(.50)
<i>CO</i>	.1850(.018)	.0205(.003)	.0034(.001)	.0170(.002)
<i>AR1</i>	.82(.04)	.66(.04)	.52(.05)	.88(.05)
<i>AR2</i>	-.14(.04)	-.20(.04)	-.19(.05)	-.09(.05)
<i>Model Var.</i>	.6372(.039)	.0489(.004)	.0098(.001)	.0082(.001)
<i>Obs. Var.</i>	.0243(.003)	.0030(.001)	.0012(<.0005)	.0013(<.0005)
<b>Interior-Valley</b>				
<i>Constant</i>	16.827	3.213	.403	2.354
<i>Temp</i>	-.182(.037)	-.033(.009)		-.025(.004)
<i>Temp</i> <sup>2</sup> × 10 <sup>4</sup>	10.30(2.3)	1.867(.6)		1.452(.30)
<i>CO</i>	.1002(.013)	.0091(.003)	.0043(.002)	.0124(.002)
<i>AR1</i>	.74(.05)	.49(.05)	.53(.05)	.75(.05)
<i>AR2</i>	-.16(.04)	-.26(.04)	-.26(.05)	-.14(.05)
<i>Model Var.</i>	.5211(.030)	.0297(.002)	.0094(.001)	.0066(.001)
<i>Obs. Var.</i>	.0304(.002)	.0023(.001)	.0010(<.0005)	.0006(<.0005)

**Table 4: Predicted Mortality Gradients**

Deaths/100,000/ppm CO/day

Section	Age	Mortality	Std. Error
<b>Central-Coastal</b>			
	65+	.185	.018
	45-64	.021	.003
	0-9	.003	.001
	Total	.017	.002
<b>Interior-Valley</b>			
	65+	.100	.013
	45-64	.009	.003
	0-9	.004	.002
	Total	.012	.002

Table 3 gives the estimated parameters for all age groups and total in both the Central-Coastal and Interior-Valley regions. Values shown in parentheses are standard errors for the estimates obtained from bootstrapping. The temperature terms were not included in the models for 0-9 age group because their coefficients were not statistically significant at .05 level.

The complete bootstrap sampling distributions for the estimators for the 65+ age group and the Central-Coastal region are shown in Figure 19. The distributions are based on 200 bootstrap replications; each one required 12 hours of computer time on a 486-33 microcomputer. It can be seen that the bootstrap sampling distributions of parameters are fairly close to normal distribution. Hence, reasonable confidence intervals for the parameters can be constructed using the normal distribution. The temperature terms were not statistically significant for the 0-9 age group so that those coefficients were not included in the model.

In addition, the estimated standard errors for the estimated parameters shown in Table 3 are all statistically significant as based on the bootstrap estimators for the standard errors. To test the hypothesis that to a parameter  $\theta$  is zero, compute

$$z = \frac{\hat{\theta}}{\hat{se}(\hat{\theta})},$$

where  $\hat{se}(\hat{\theta})$  denotes the bootstrap estimate of the standard error, and compare the result to percentage points of a standard normal variable. For example, to test the coefficient for CO level, compute  $\frac{.185}{.018} = 10.3$  which is highly significant. It is clear, using this procedure, that we strongly reject the hypothesis that any parameter is zero in the Central-Coastal over-65 age group. To get a 95% confidence interval, use the formula

$$\hat{\theta} \pm 1.96\hat{se}(\hat{\theta});$$

which yields the interval .150 to .220 for the gradient effect of CO level on mortality, expressed as deaths per 100,000 per day.

For purposes of summarizing the information conveyed by these tables, consider Table, 4 which shows only the effect of the CO level on mortality for each of the age groups and total in each region. The interpretation of these numbers is that each of these numbers is the predicted increment in deaths per 100,000 per day when the temperature is held fixed and the CO level increases by 1 ppm. For example, the .185 deaths per 100,000 per ppm CO per day translates to about 68 deaths per year per additional ppm CO. One should

include the usual precautionary note that integrating increased rates over a considerable number of days may be dangerous, since it is plausible that deaths that occur on high pollution days might simply have been postponed if the pollution had been lower. On examination of this table, one is struck again by the differences in mortality between the two regions and among the different age groups. It is clear on comparing again with Table 2 that the levels are proportional to the differences in overall mortality rates between the two regions, which may be attributable to different racial compositions of the regions populations. We may draw the general conclusion that a higher death rate for a subgroup implies that they will incur a higher incremental death rate due to pollution.

### 4.3 Parametric Mortality Prediction

Using *Model II*, which is a better parametric fit for the data, an heuristically reasonable predicted mean mortality as a function of temperature and pollution is

$$\hat{M}_t = \hat{a} + \hat{c}_1 T_t + \hat{c}_2 T_t^2 + \hat{d} P_t$$

where  $\hat{a}$  accumulates the average effects of detrending the series. Note that *Model II* was fitted to the detrended time series.

Tables 20 (a-d) show the result of substituting into the above prediction equation the parameters from Table 3 as a function of temperature and CO level only. That is, these give the predicted mortalities in the absence of the seasonal mortality signal. The contours can be used to read off predicted daily mortality values at various combinations of temperature and CO level. The predicted contours for the Central-Coastal region as shown in Figure 20(a,b) match up quite well with the values predicted using the nonparametric smoothing approach displayed in Figure 15(a-d). Predicted contours for the Interior-Valley region also match up quite well, as can be seen by examining Figure 20(c,d) (compare the nonparametric contours in Figure 16(a-d)).

Based on agreement of parametric and nonparametric analyses, we are fairly certain that there is an underlying unobservable mortality signal that can be isolated. Isolating this signal not change the predicted mortality contours as a function of temperature and pollution, so these quantities are still contributing to the explanation of gradients in mortality. Furthermore, the nonparametric profiles do not seem to be biased by this underlying mortality signal. Possibly its effects are smoothed out by the heavy averaging over the pollution and temperature profiles. For example, it is plausible that the same pollution

and temperature values occur on many different amplitude locations of the mortality signal so that averaging over these different locations essentially smoothes out the overall effect.

## 5. SUMMARY AND CONCLUSIONS

The primary objective of this study has been to develop a set of procedures for predicting mortality as a function of weather and pollution effects in Los Angeles County by age and by region of residence. We developed both nonparametric and parametric procedures for this purpose.

The nonparametric approach involved simply smoothing mortality over common temperature and pollution coordinates to develop mortality profiles for prediction as a function of temperature and pollution level. Using this approach we were able to reconfirm the quadratic variations in mortality with temperature. This means that death rates are higher on both warmer and colder days and that the mean gradient is predictable as a function of temperature. The nonparametric contours also confirm a linear or possibly loglinear variation of mortality with pollution.

The primary purpose of the parametric approach was to see whether mortality fluctuations could be modeled solely by a linear time trend and an unobservable cyclical mortality component which explained both the observed mortality and its underlying time correlation structure. Estimating this component in the absence and presence of covarying pollution and temperature effects (*Models I and II, respectively*) showed that *Model II*, which includes the covariates, fits the data significantly better. We also showed that there was a significant unobservable mortality signal which was extracted and examined. The nonlinear profiles predicted by the parametric model with the unobserved component and trend accounted for were reasonably consistent with those produced by the nonparametric procedure. The statistical analysis of the parametric structural model required developing a parameter estimation and smoothing methodology for the mixed linear dynamic model.

Specifically, we found that mean predicted mortality in the most susceptible age group 65 and over, changes by .185 deaths per 100,000 per day per ppm CO in the Central-Coastal region and by only .100 deaths per 100,000 per day per ppm CO in the Interior-Valley region; this difference was statistically significant. Analogous results for the 45-64 age groups were .021 for the Central-Coastal region and .004 for the Interior-Valley region. For the 0-9 age group the rates were nearly the same with .003 and .004 deaths respectively in the Central-Coastal and Interior-Valley groups. Generally speaking, both the population death rates per 100,000 and the changes in those rates induced by varying temperature

and pollution levels were higher in the Central-Coastal region than in the Interior-Valley region. It is probable that these large differences are due to the differences in overall death rates (see Table 2); such differences in overall death rates can probably be explained by the different racial compositions of the two regions' population.

From the standpoint of statistical methodology, the results obtained in the study indicate that applying nearest neighbor smoothing procedures leads to satisfactory predicted mean mortality profiles that are not adversely affected by trend, time correlation, or the presence of unobservable random components. Furthermore, these profiles are very helpful for the process of developing reasonable functional forms for parametric models. The parametric method, however, yields coefficients that are more easily interpreted for practical epidemiological work, since they predict the mean mortality change directly as a known function of temperature and pollution levels. Furthermore, the parametric analyses can include reasonable standard error estimates computed by using the recently developed bootstrap methodology of Stoffer and Wall (1991).

## 6. References

- Cass G.R. and M.H. Conklin (1984) Elemental carbon concentrations: Estimation of an historical data base. *Atmospheric Environment*, **18**, 153-162.
- Conklin M.H., G.R. Cass, L.-C. Chu and E.S. Macias (1982). Wintertime carbonaceous aerosols in Los Angeles - An exploration of the role of elemental carbon. In *Atmospheric Aerosol: Source/Air Quality Relationships* (edited by Macias E.S. and Hopke P.K.), American Chemical Society. Washington D.C.
- Fairley, D. (1990). The relationship of daily mortality to suspended particulate in Santa Clara County, 1980-1986. *Env. Health Perspectives*, **89**, 159-168.
- Gray H.A. and G.R. Cass (1986). Characteristics of atmospheric organic and elemental carbon particle concentrations in Los Angeles. *Environmental Science and Technology*, **20**, 580-588.
- Hechter, H.H. and J.R. Goldsmith (1961). Air pollution and daily mortality. *Am. J. Med. Sci*, **24**, 581-588.
- Hexter, A.C. and J.R. Goldsmith (1971). Carbon monoxide: Association of community air pollution with mortality. *Science*, **172**, 265-266.
- Kinney, P.L. and H. Özkaynak (1991). Associations of daily mortality and air pollution in Los Angeles County. *Environ. Res.*, **54**, 99-120.
- Lebowitz, M.D. (1973). A comparative analysis of the stimulus-response relationship between mortality and air pollution- weather. *Environ. Res.*, **6**, 106-118.
- Martin, A.E. and W.H. Bradley (1960). Mortality, fog and atmospheric pollution: An investigation during the winter of 1958- 59. *Mon. Bull. Minist. Health*, Public Health Service Lab., Series **19**, 56-72.
- Mazumdar, S., H. Schimmel and I.T.T. Higgins. Relationship of daily mortality to air pollution; An analysis of 14 London winters 1958/59-1971/72. *Arch. Environ. Health*, **37**, 213-220.
- Mills, C.A. (1960). Respiratory and cardiac deaths in Los Angeles smogs during 1956, 1957 and 1958. *Am. J. Med. Sci.*, **239**, 307-315.

- Newton, H.J., G.R. North and T.J. Crowley (1991). Forecasting Global Ice Volume. *J. Time Series Analysis*, **12**, 255-265.
- Oechsli, F.W. and R.W. Buechley (1970). Excess mortality associated with the Los Angeles September hot spells. *Environ. Res.*, **3**, 277-284.
- Ostro, B. (1984). A search for a threshold in the relationship of air pollution to mortality: A reanalysis of data on London winters. *Environ. Health Persp.*, **58**, 397-399.
- Schimmel, H. (1978). Evidence for possible acute health effects of ambient air pollution from time series analysis: Methodological questions and some new results based on New York City daily mortality. 1973-1976. *Bull. NY Acad. Med.*, **54**, 1052-1108.
- Schwartz, J. and A. Marcus (1990). Mortality and air pollution in London: A time series analysis. *Am. J. Epidemiol.*, **131**, 185-194.
- Shikiya, D. M. Hoggan and A. Davidson (1980). Summary of Air Quality in the South Coast Air Basin of California, 1979. South Coast Air Quality Management District.
- Shumway, R.H. (1988). *Applied Statistical Time Series Analysis*. Englewood Cliffs: Prentice Hall.
- Shumway, R.H., R. Tai, L. Tai and Y. Pawitan (1983). Statistical analysis of daily London mortality and associated weather and pollution effects. *Technical Report No. 53*, Technical Report Series, Division of Statistics, University of California, Davis 95616.
- Shumway, R.H., A.S. Azari and Y. Pawitan (1988). Modeling mortality fluctuations in Los Angeles as functions of pollution and weather effects. *Environmental Research*, **45**, 224-241.
- Shumway, R.H. and M.J. Katzoff (1991). Adjustment of provisional mortality series: The dynamic linear model with structured measurement errors. *J. Amer. Statist. Assoc.*, **86**, 611-617.
- Shumway, R.H. and A.S. Azari (1991). Non linear modeling of longitudinal mortality series: To live and die in LA. Presented at the 1991 Joint Statistical Meeting, Atlanta, GA, August 18-22.
- Stoffer, D.S. and K.D. Wall (1991). Bootstrapping state-space models: Gaussian maximum likelihood estimation and the Kalman filter. *J. Amer. Statist. Assoc.*, **86**, 1024-1033.

Wakamatsu, P. (1991). Personal Communication. Data Collection and Analysis, Los Angeles Department of Health Services, 313 North Figuera St., Los Angeles, CA 90012.

Ware, J.H., L.A. Thibodgan, F.E. Speizer, S. Colome and B.J. Ferris (1981). Assessment of the health effects of atmospheric sulfur oxides in particulate matter: Evidenc from observational studies. *Env. Health Perspectives*, **41**, 256-275.

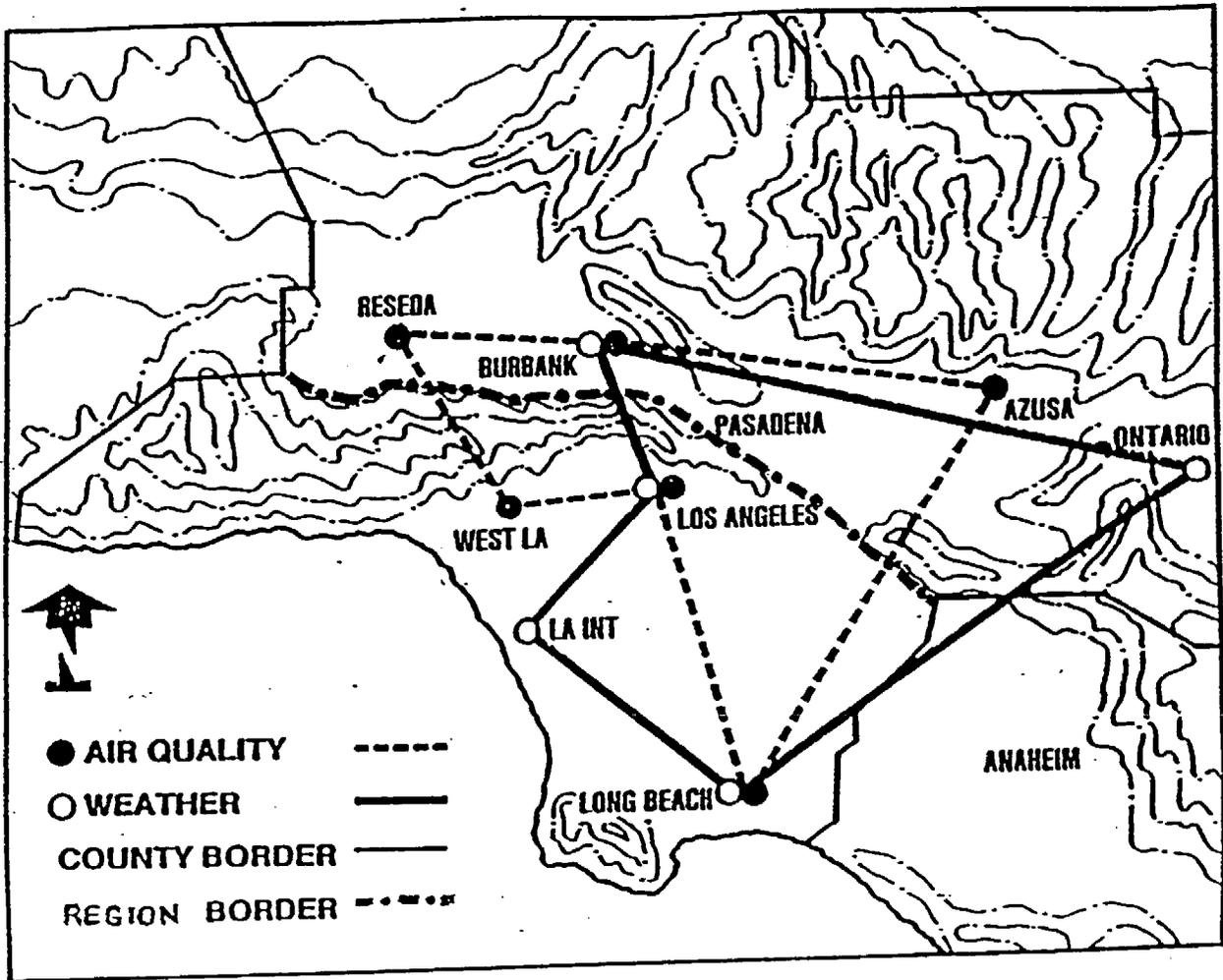


Figure 1: Map of Los Angeles County Partitioned into Central-Coastal and Interior-Valley Sections with Air Quality and Weather Stations.

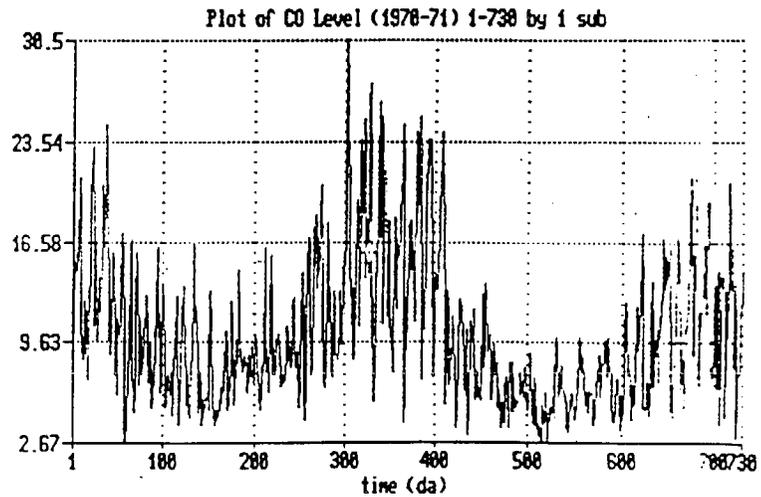
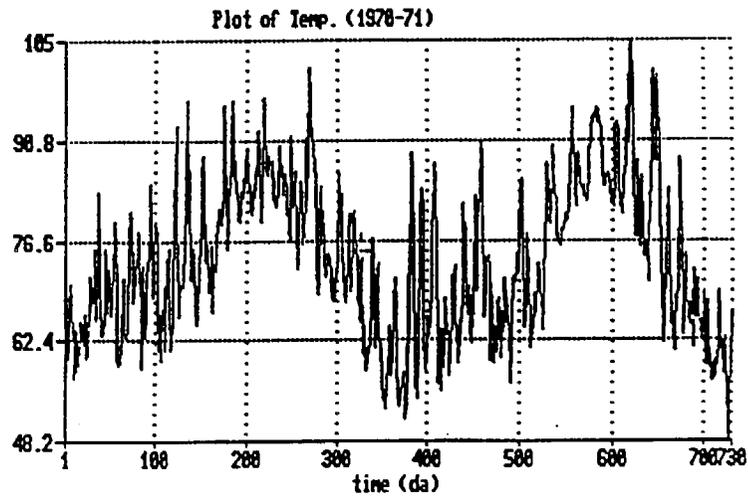
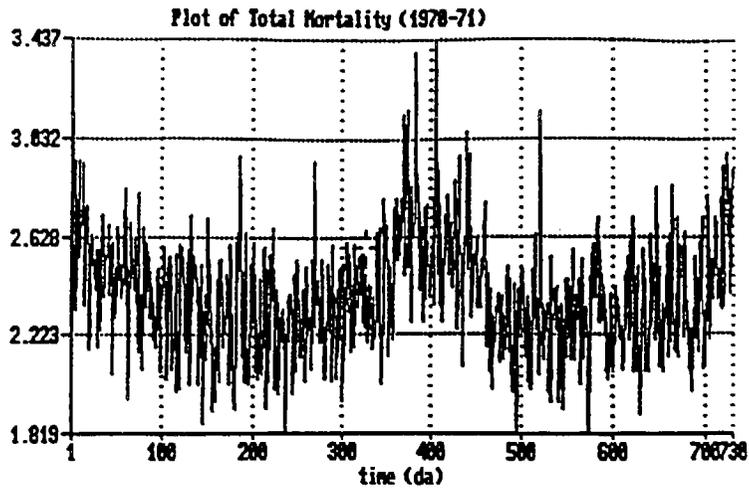


Figure 2: Original Daily Records for Total Mortality, Temperature and Carbon Monoxide (CO) Level (1970-1971).

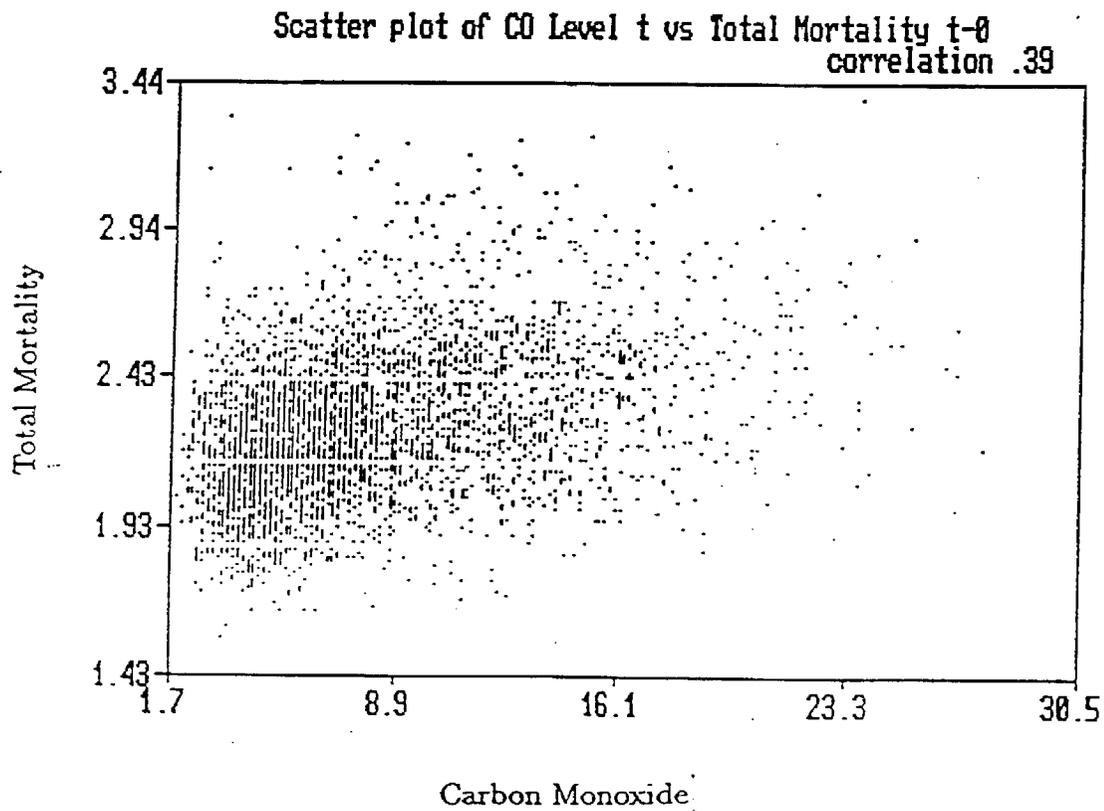
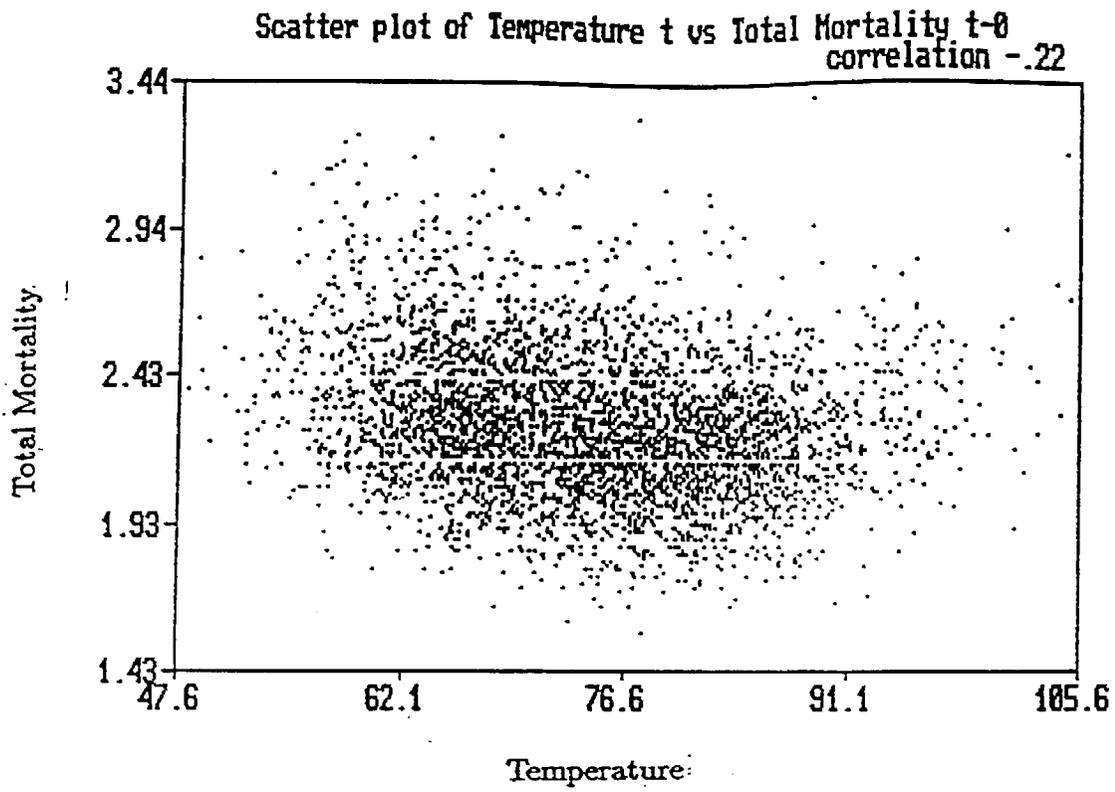
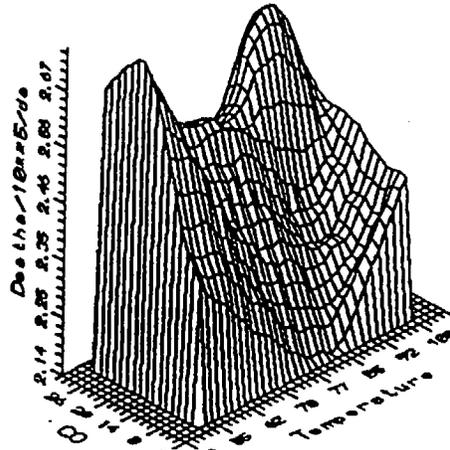


Figure 3: Scatterplots Relating Temperature and CO to Total Mortality (1970-1979, 3652 days).



Total Mort. vs Temp. and CO (LA County)

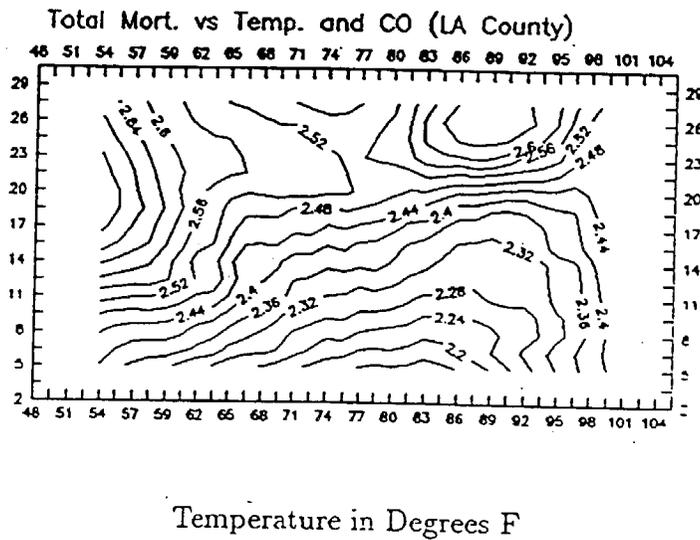


Figure 4: Nonparametric Surface and Contour Plots Relating Total Mortality to Temperature and CO Level (1970-1979, 3652 days).

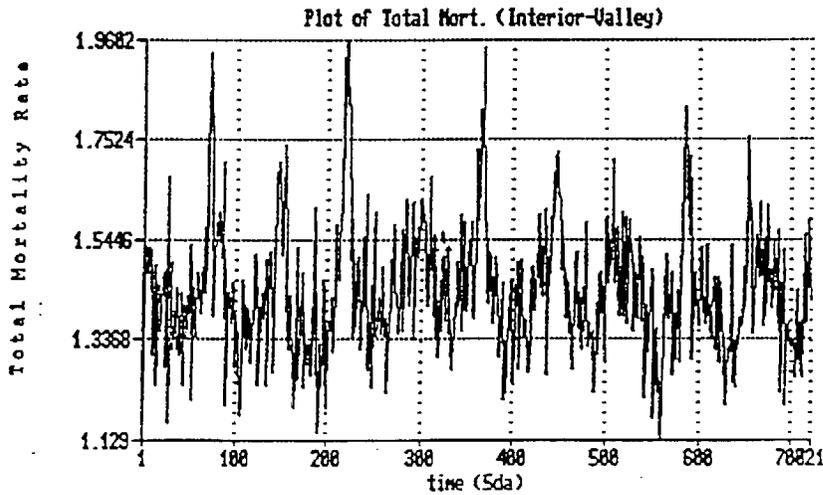
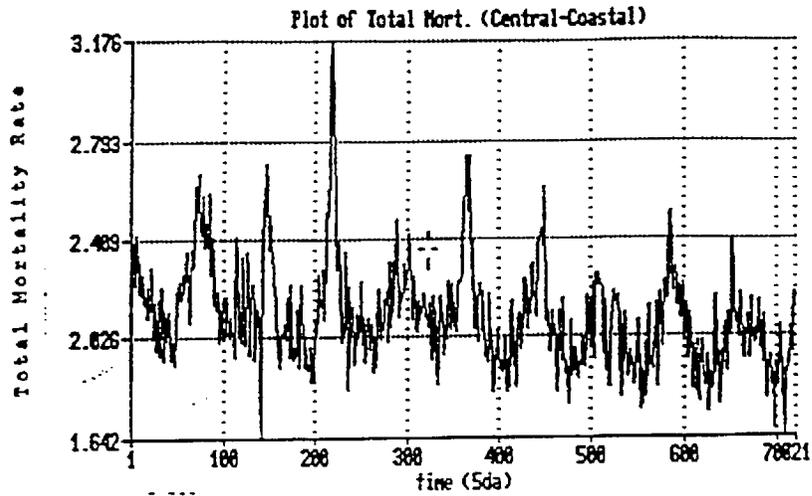


Figure 5: Total Mortality Rate (Deaths/100,000/day) for Central-Coastal and Interior-Valley Sections of Los Angeles Co. (1970-1979, 721 5-day intervals).

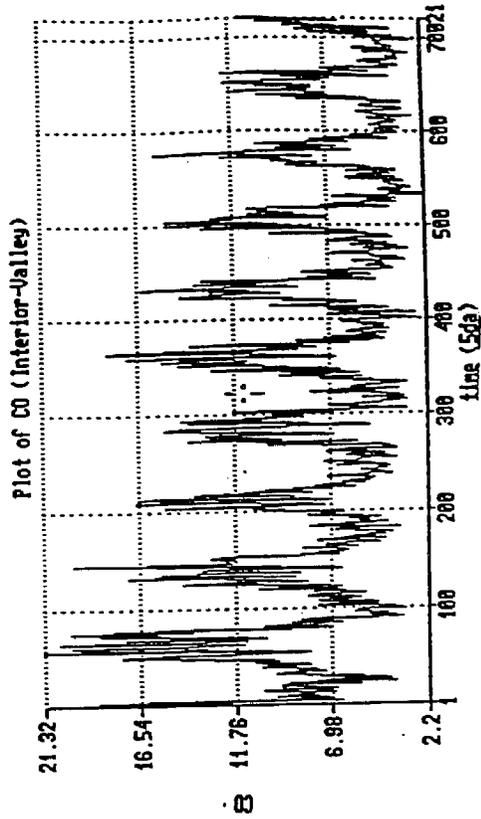
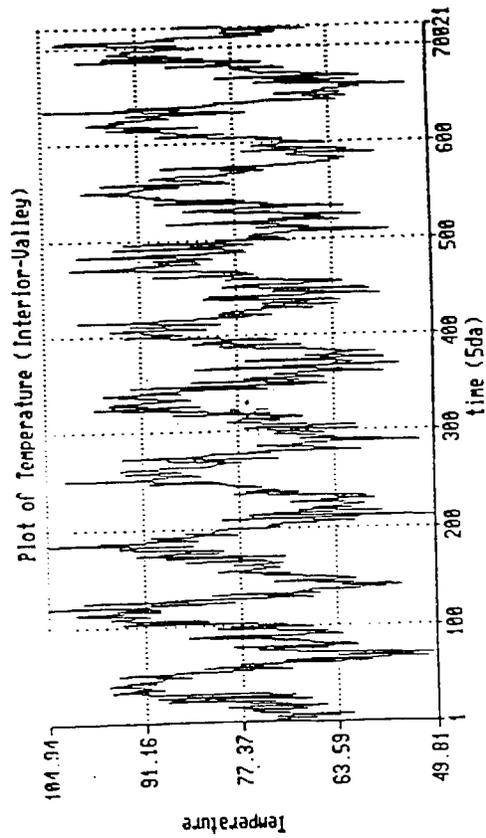
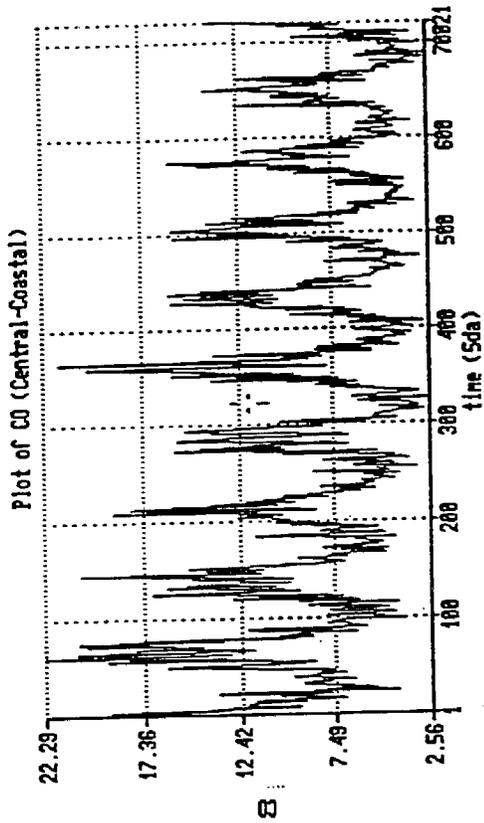
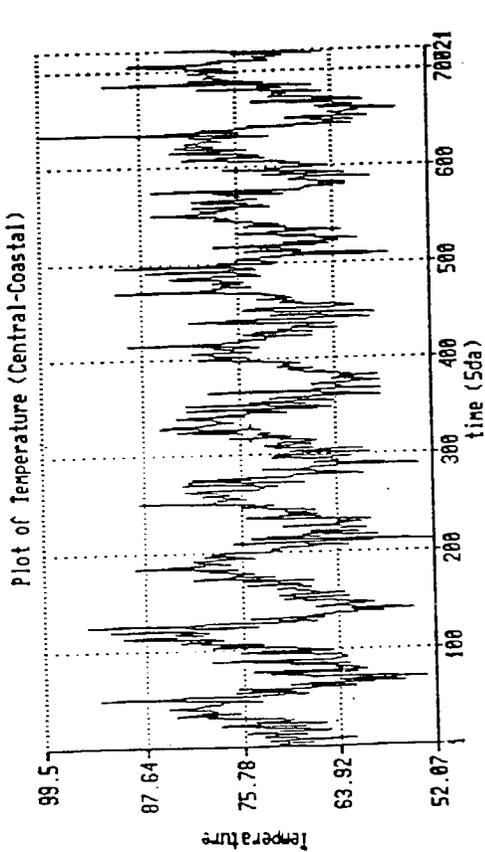
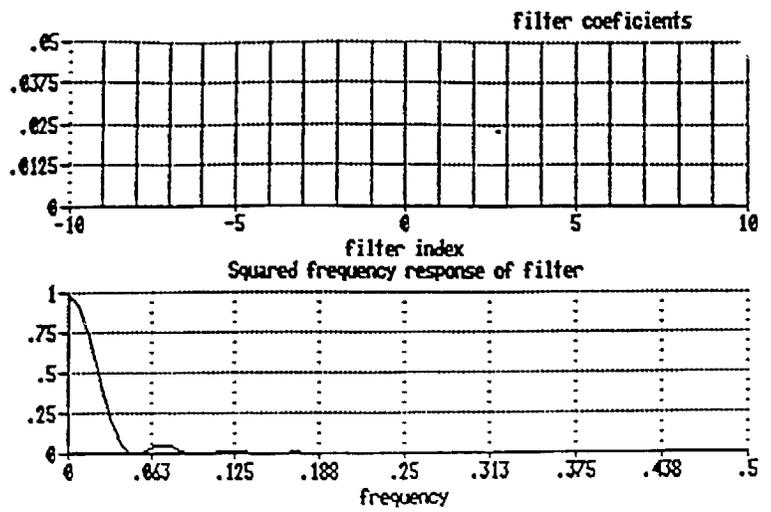


Figure 6: Temperature and CO Levels for Central-Coastal and Interior-Valley Sections of Los Angeles Co. (1970-1979, 721 5-day intervals).



.0123081  
 .0200389  
 .0295593  
 .0403846  
 .0518010  
 .0629365  
 .0728599  
 .0806902  
 .0857055  
 .0874320  
 .0857055  
 .0806902  
 .0728599  
 .0629365  
 .0518010  
 .0403846  
 .0295593  
 .0200389  
 .0123081

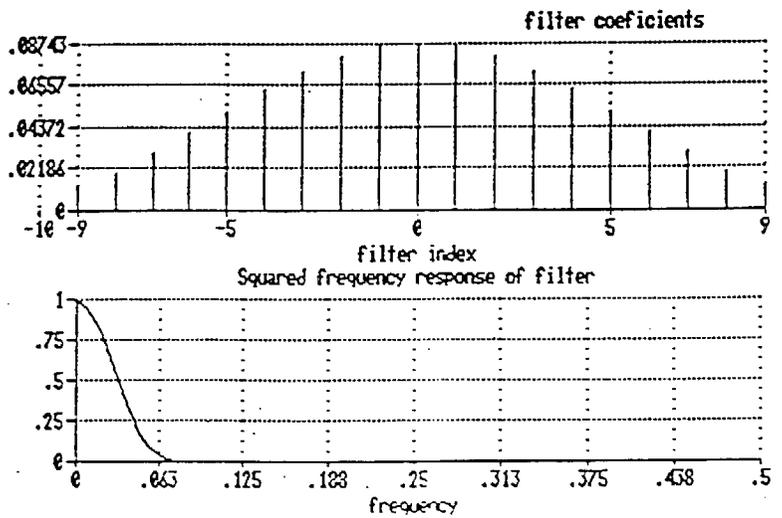


Figure 7a: Filters Applied to Los Angeles Data Are 20 pt. Moving Average (upper panel) and Linear Filter Designed by Shumway et al (1988)(lower panel).

0  
 7.857951E-04  
 -5.806446E-04  
 -2.252512E-03  
 -3.569921E-03  
 -3.846702E-03  
 -2.610336E-03  
 1.852824E-04  
 4.045931E-03  
 7.973859E-03  
 1.066942E-02  
 1.086566E-02  
 7.721509E-03  
 1.172646E-03  
 -7.862573E-03  
 -1.749903E-02  
 -2.515068E-02  
 -2.801328E-02  
 -2.367303E-02  
 -1.070955E-02  
 1.084886E-02  
 3.933661E-02  
 .0715939  
 .1034346  
 .1303573  
 .1483597  
 .1546875  
 .1483597  
 .1303573  
 .1034346  
 .0715939  
 3.933662E-02  
 1.084888E-02  
 -1.070954E-02  
 -2.367303E-02  
 -2.801327E-02  
 -2.515067E-02  
 -1.749903E-02  
 -7.86257E-03  
 1.17265E-03  
 7.721511E-03  
 1.086566E-02  
 1.066942E-02  
 7.973857E-03  
 4.045931E-03  
 1.852842E-04  
 -2.610334E-03  
 -.0038467  
 -3.569915E-03  
 -2.252507E-03  
 -5.806434E-04

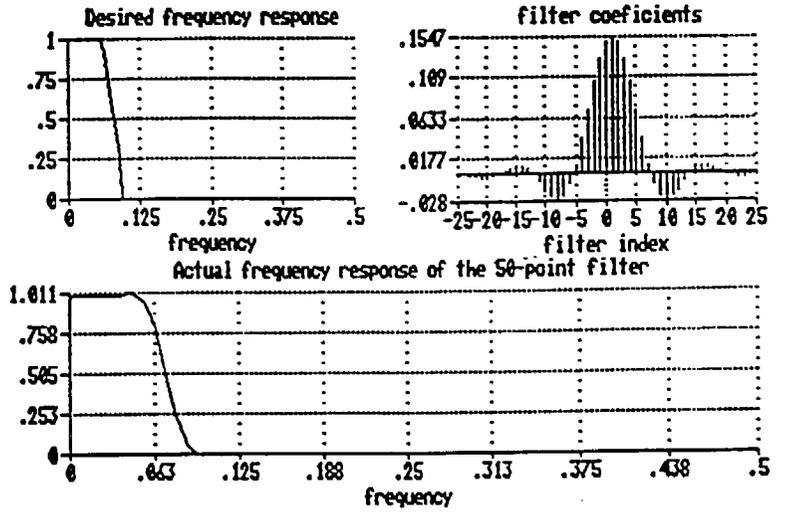


Figure 7b: Filters Applied to Los Angeles Data in This Study.

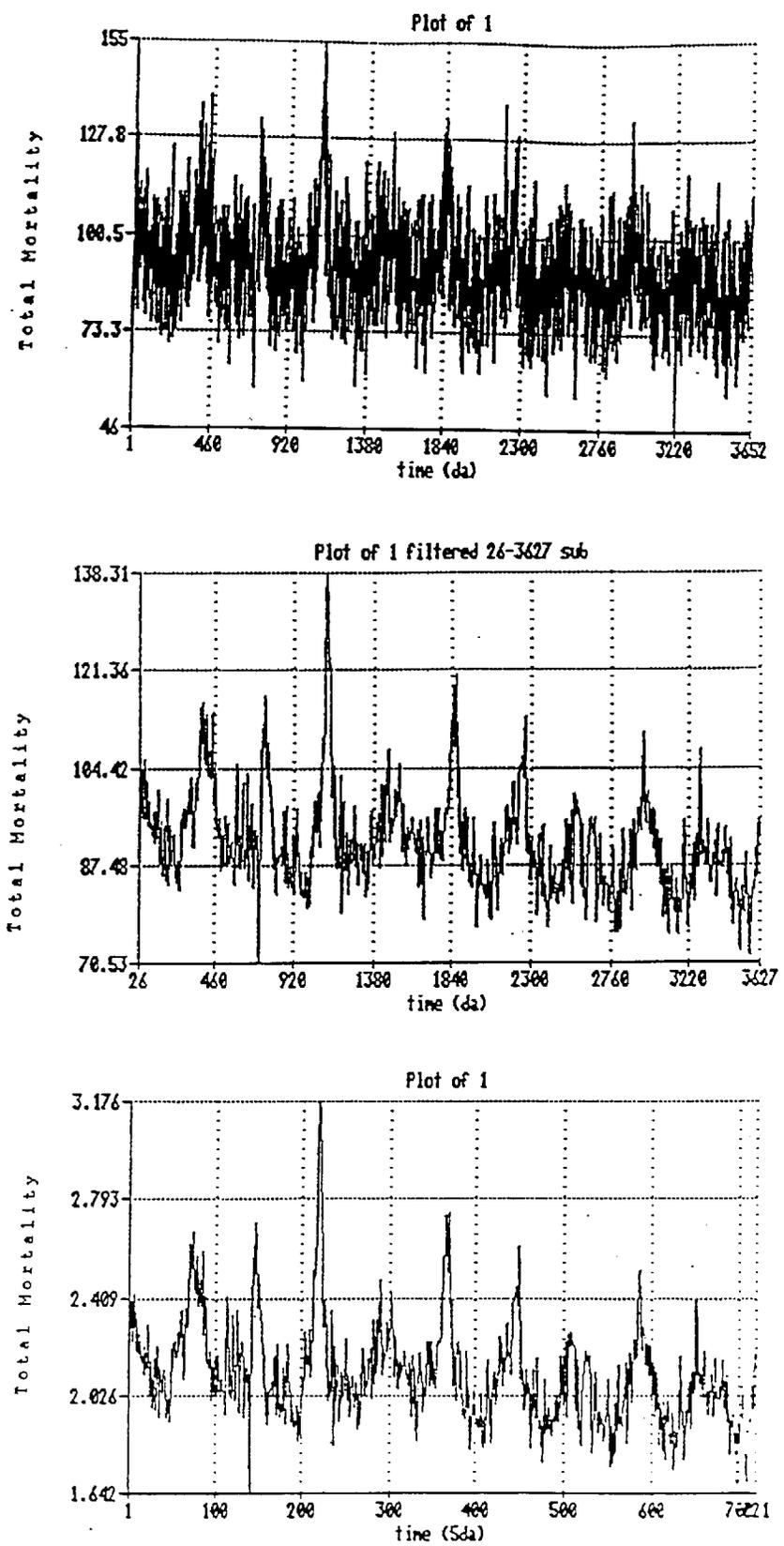


Figure 8: Raw, Filtered and Subsampled Filtered Central-Coastal Total Mortality Record (1970-1979, 721 5-day intervals).

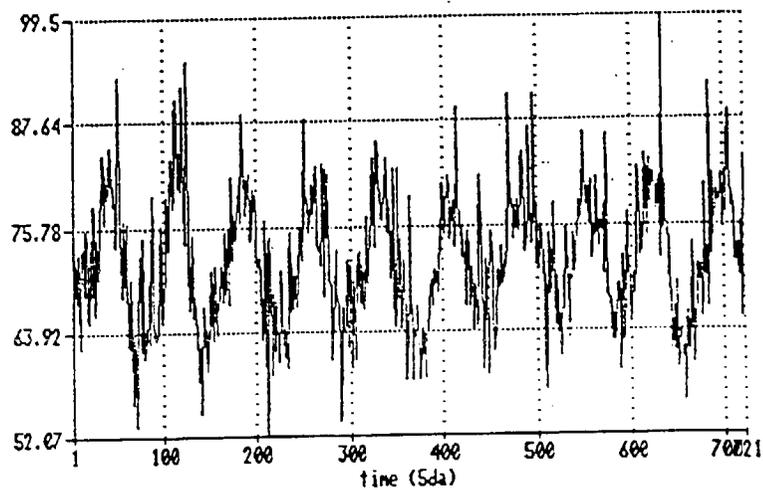
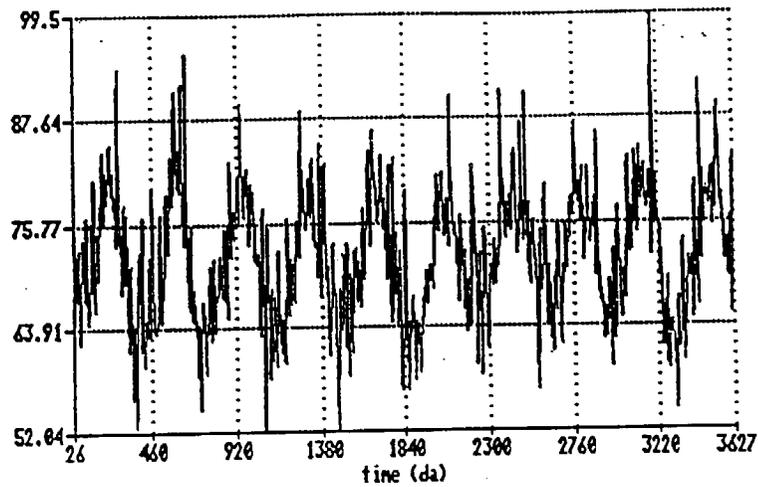
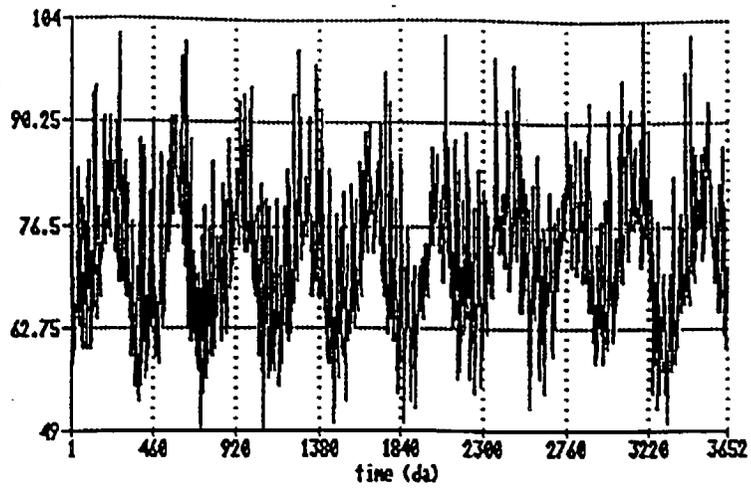


Figure 9: Raw, Filtered and Subsampled Filtered Central-Coastal Temperature Record (1970-1979, 721 5-day intervals).

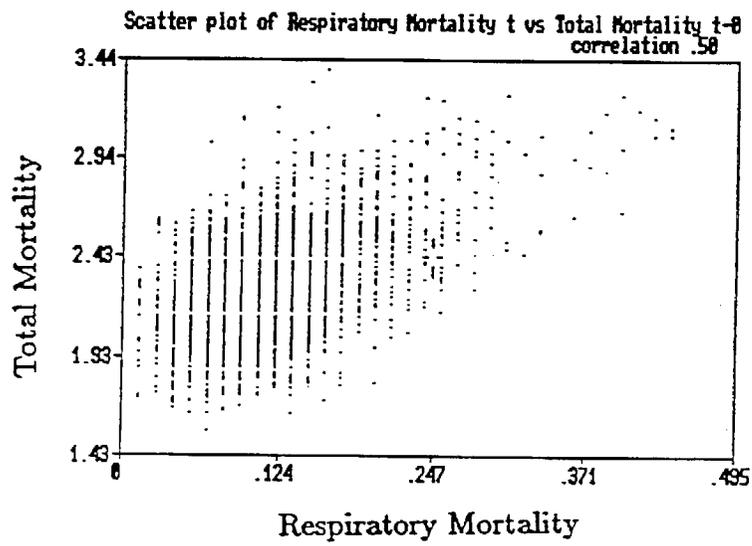
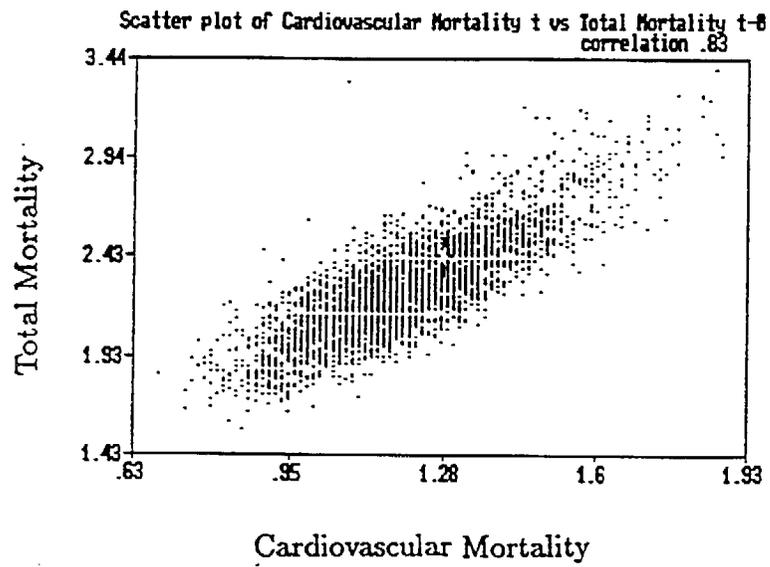


Figure 10: Scatterplots Relating Respiratory and Cardiovascular Mortality to Total Mortality (1970-1979, 3652 days).

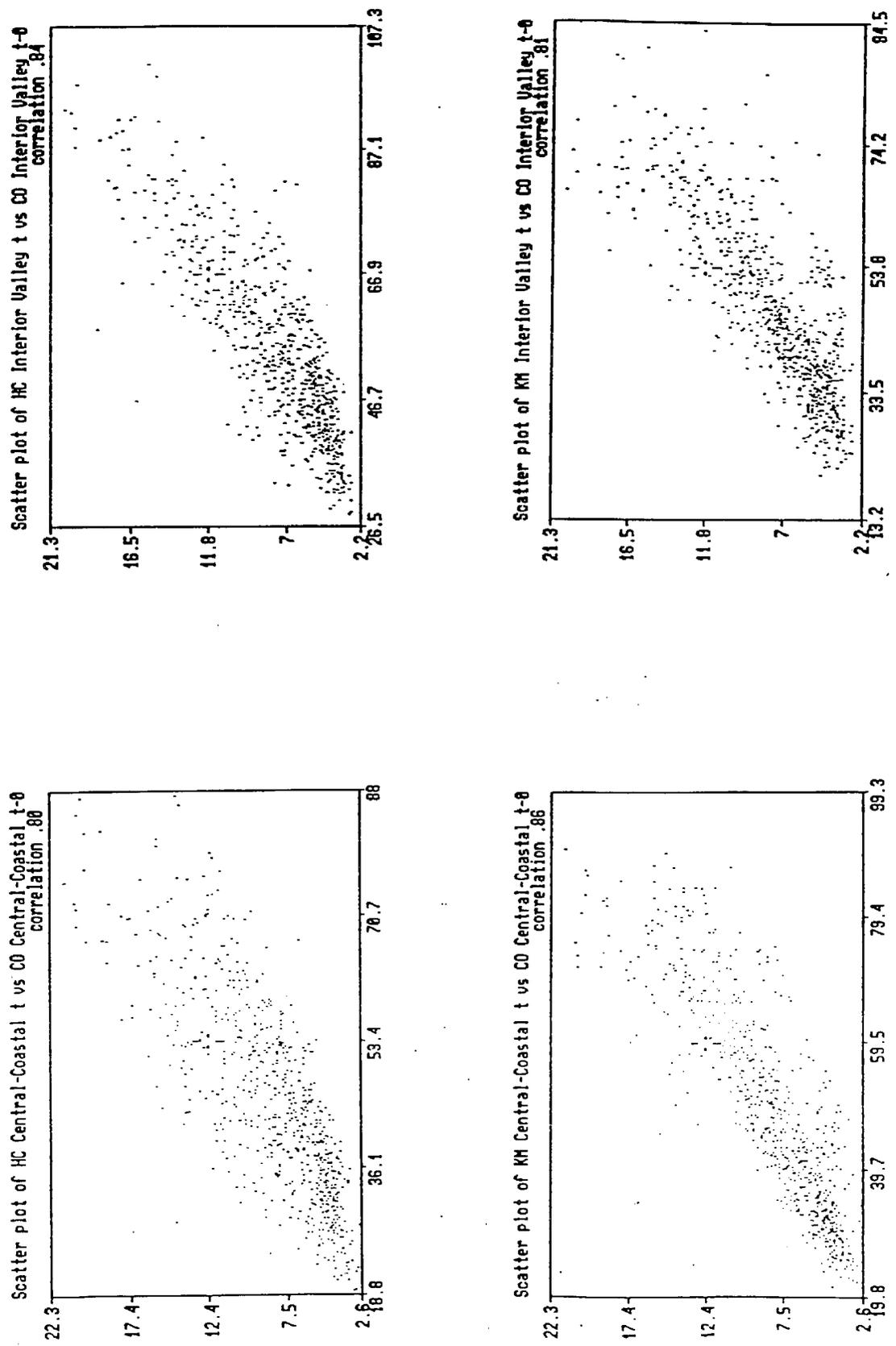


Figure 11: Scatterplots Relating CO Used in Study to Other Correlated Pollutants HC and KM for Central-Coastal and Interior Valley Sections (1970-1979, 721 5-day intervals).

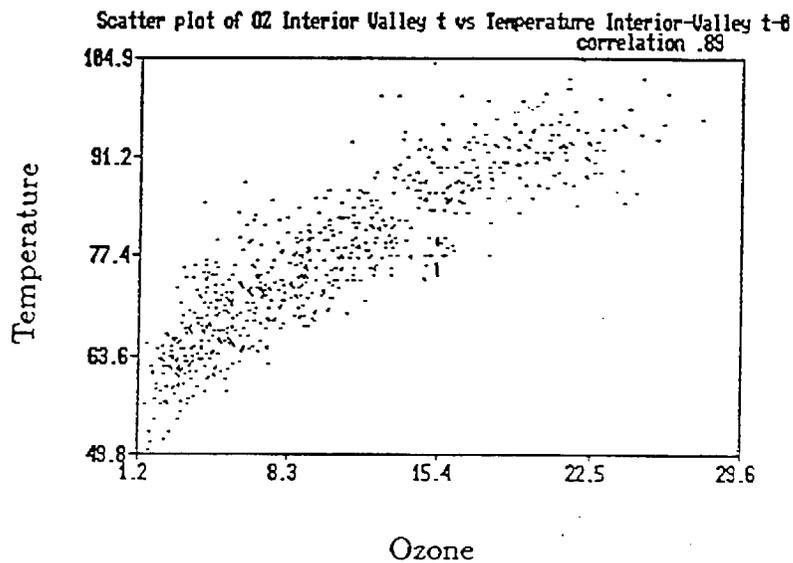
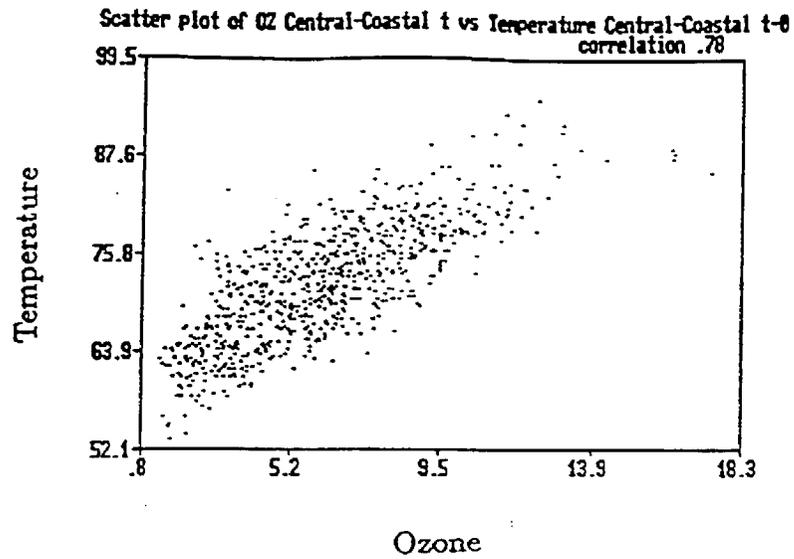
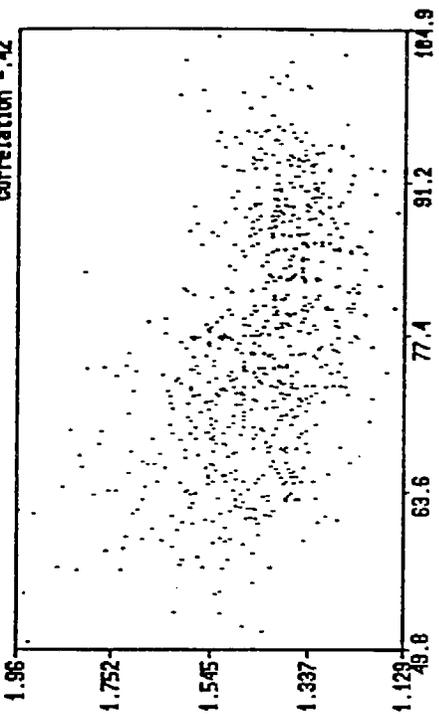
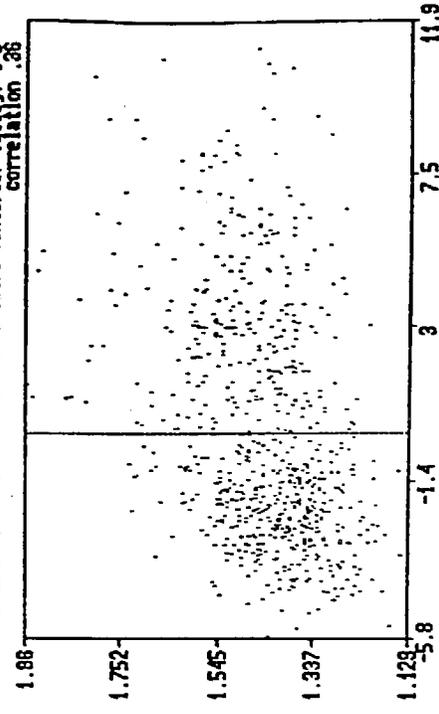


Figure 12: Scatterplots Relating Ozone and Temperature, for Central-Coastal and Interior-Valley Sections in Los Angeles Co. (1970-1979, 721 5-day intervals).

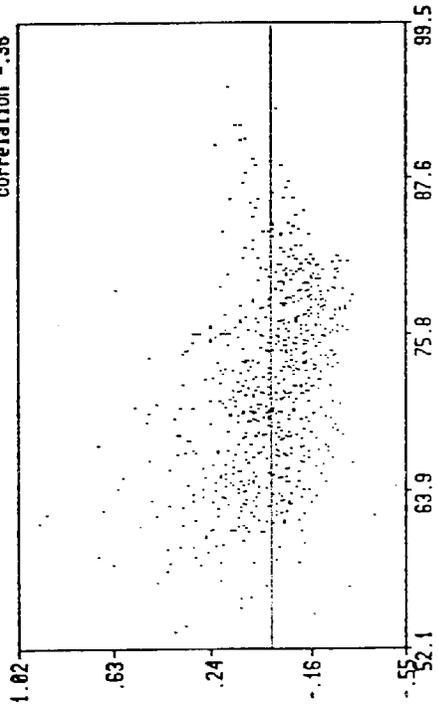
Scatter plot of Temperature (Interior-Valley) t vs Total (Interior-Valley) t-0  
correlation = .42



Scatter plot of CO linear t vs Total (Interior-Valley) t-0  
correlation = .36



plot of Temperature (Central-Coastal) t vs Total Mort. (Central-Coastal) linear  
correlation = .36



plot of CO (Central Coastal) linear t vs Total Mort. (Central-Coastal) linear  
correlation = .48

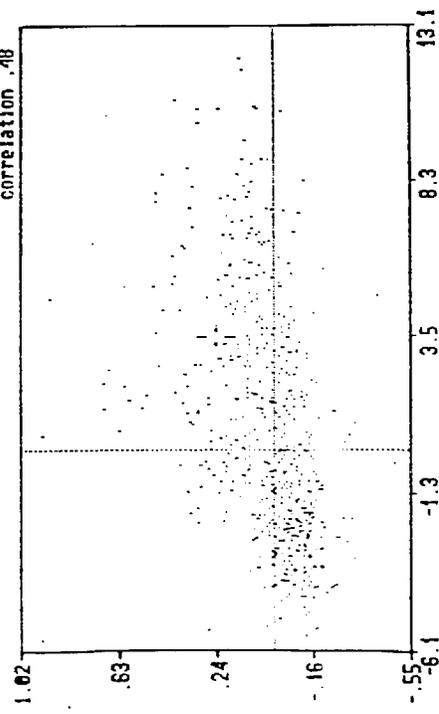


Figure 13: Scatterplots Relating Temperature and Detrended CO to Detrended Mortality for Central-Coastal and Interior-Valley Sections. (1970-1979, 721 5-day intervals).

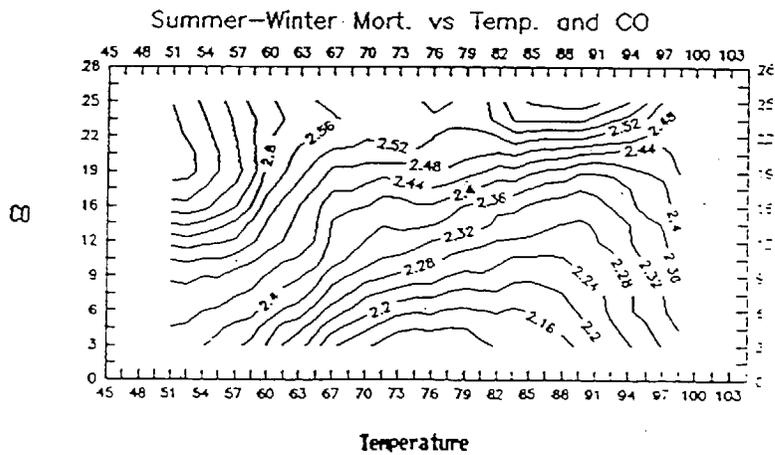
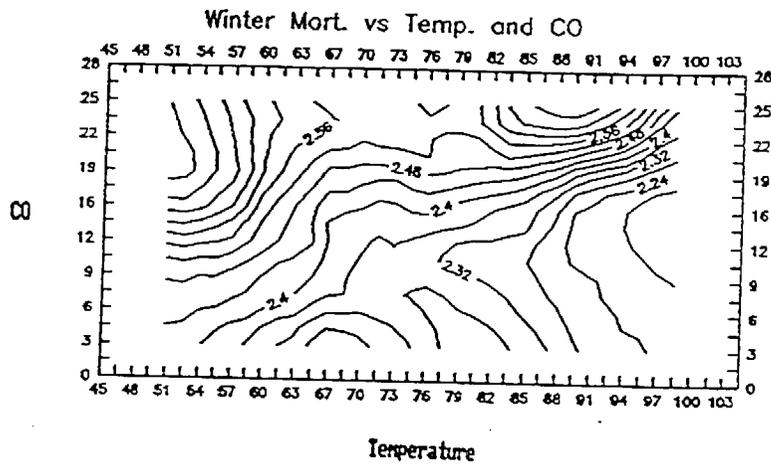
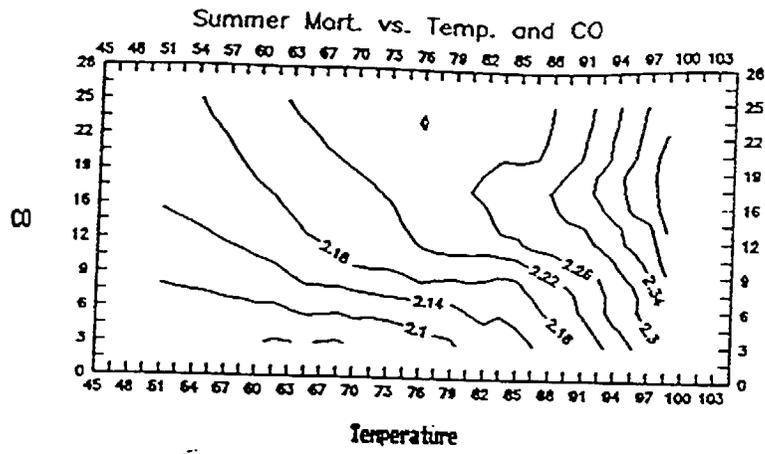
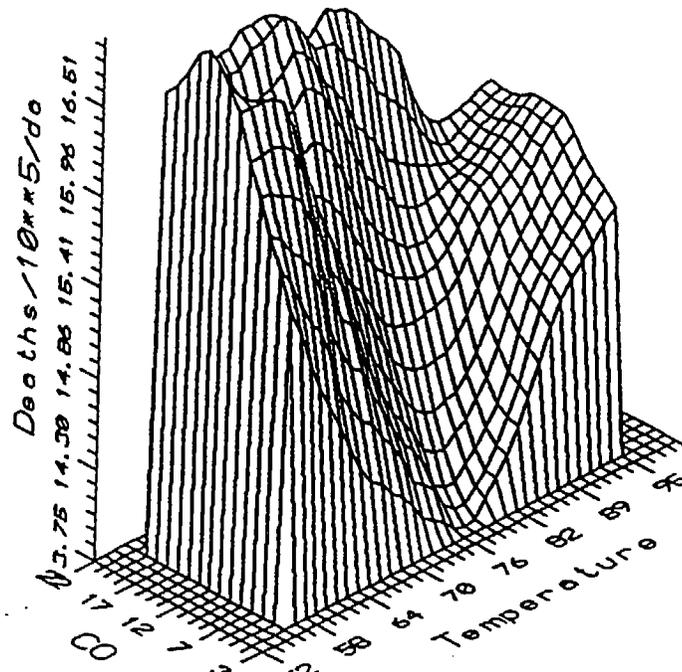


Figure 14: Comparison of Summer and Winter Mortality Profiles with Temperature and CO (1970-1979, Raw Data, 128 days per Season).



Mort. 65+ vs Temp. and CO (Central-Coastal)

Mort. 65+ vs Temp. and CO (Central-Coastal)

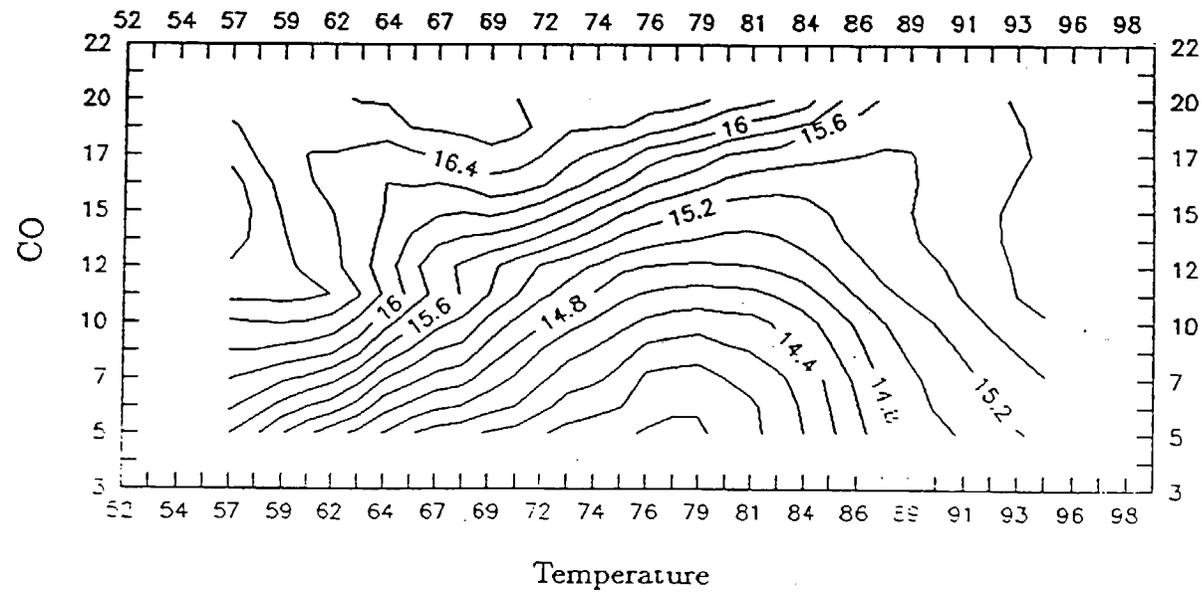
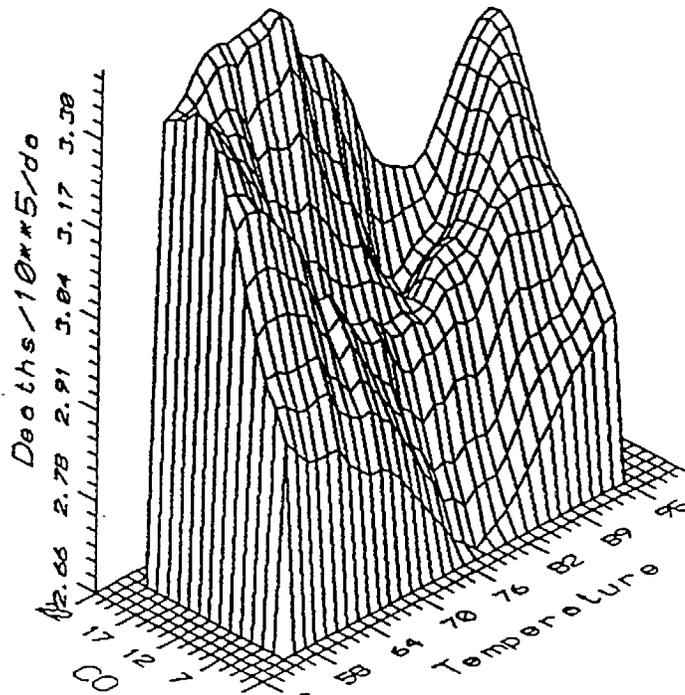


Figure 15a: Mortality Rate (deaths/100,000/day) Profiles and Contours With Temperature and CO for Age 65+ for Central-Coastal Section (1970-1979, 721 5-day intervals).



Mort. 45-64 vs Temp. and CO (Central-Coastal)

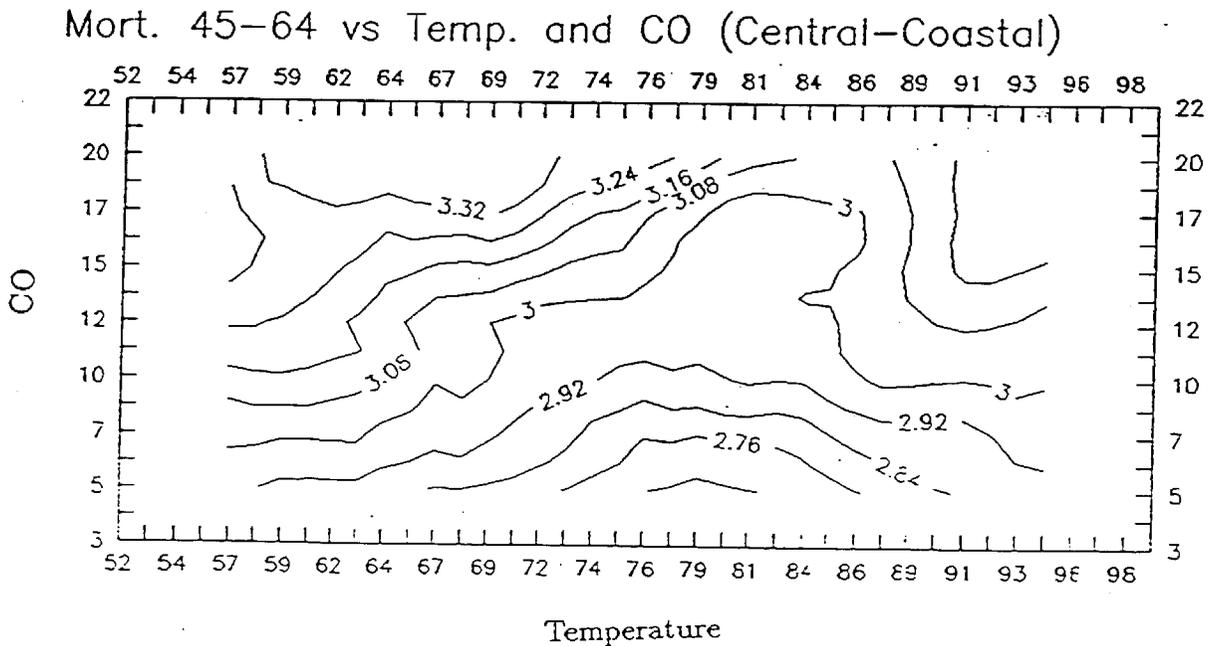
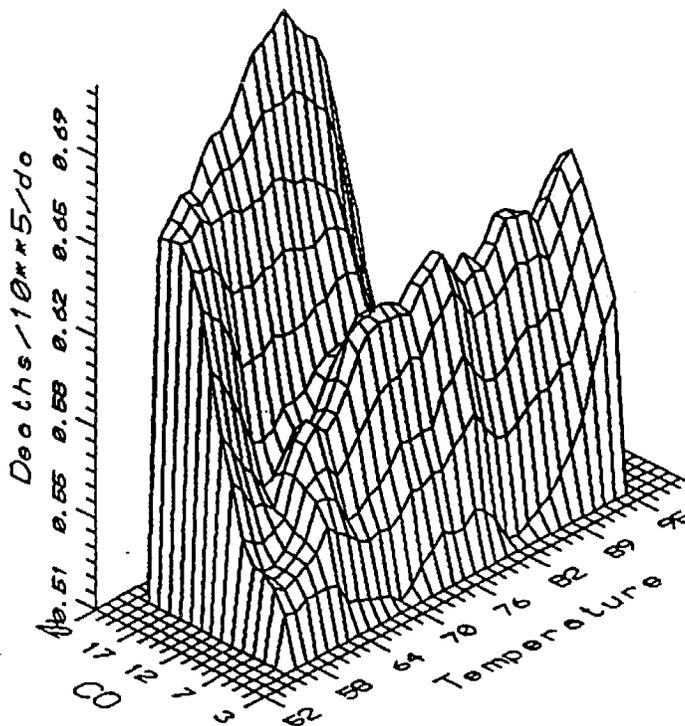


Figure 15b: Mortality Rate (deaths/100,000/day) Profiles and Contours With Temperature and CO for Age 45-64 for Central-Coastal Section (1970-1979, 721 5-day intervals).



Mort. 0-10 vs Temp. and CO (Central-Coastal)

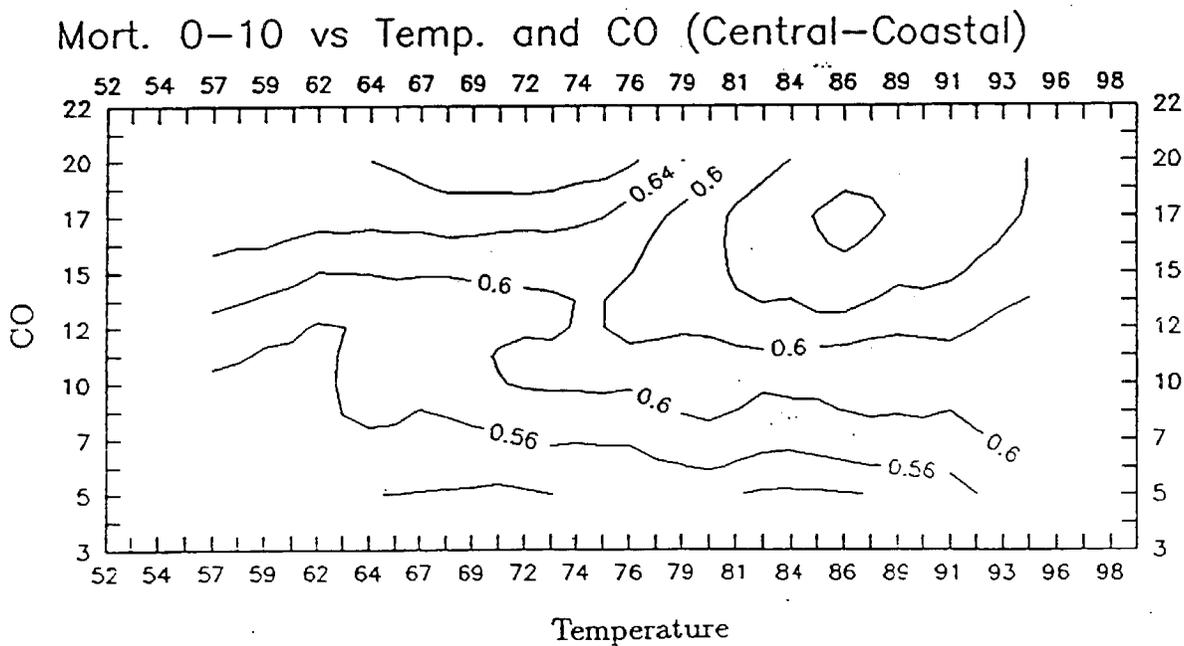
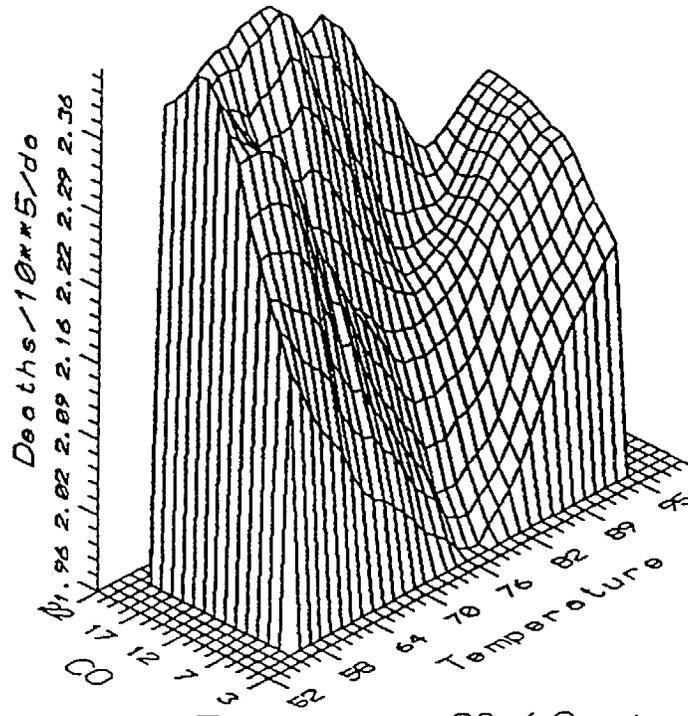


Figure 15c: Mortality Rate (deaths/100,000/day) Profiles and Contours With Temperature and CO for Age 0-9 for Central-Coastal Section (1970-1979, 721 5-day intervals).



Total Mort. vs Temp. and CO (Central-Coastal)

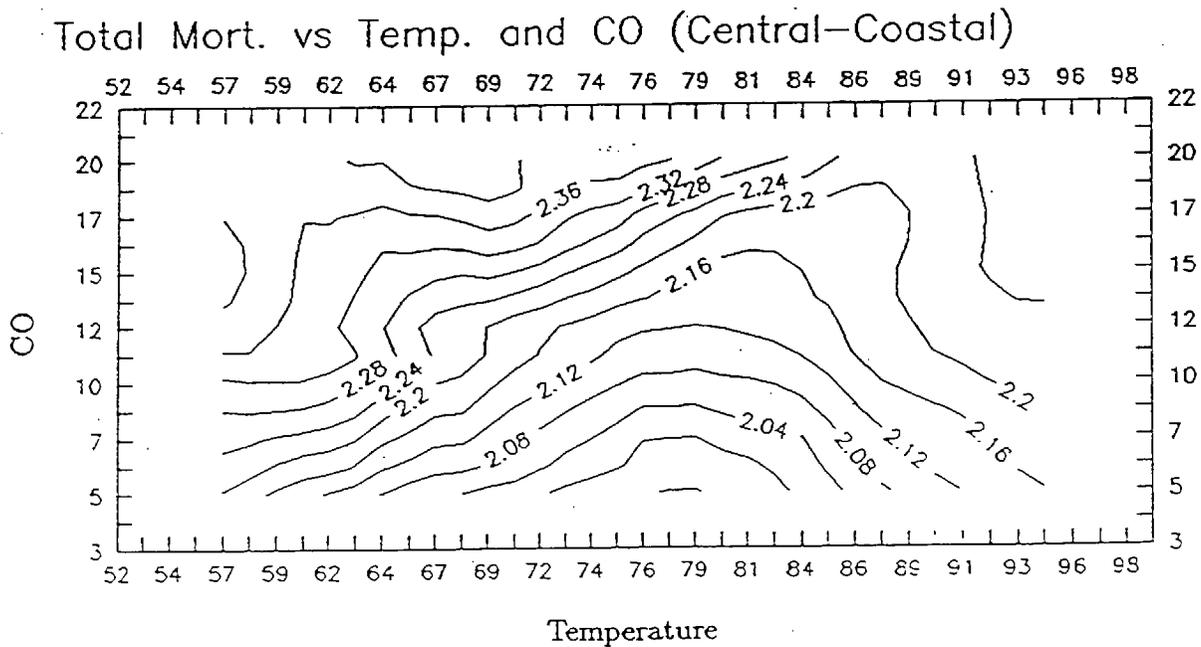
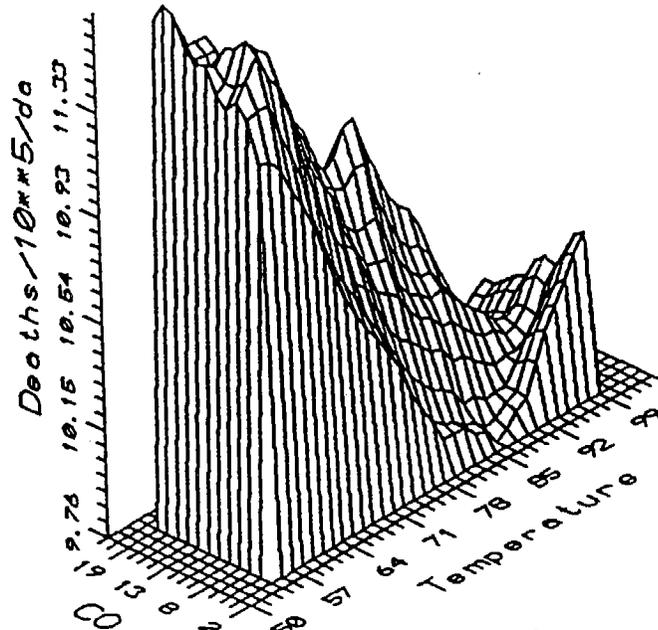


Figure 15d: Mortality Rate (deaths/100,000/day) Profiles and Contours With Temperature and CO (Total) for Central-Coastal Section (1970-1979, 721 5-day intervals).



Mort. 65+ vs Temp. and CO (Interior-Valley)

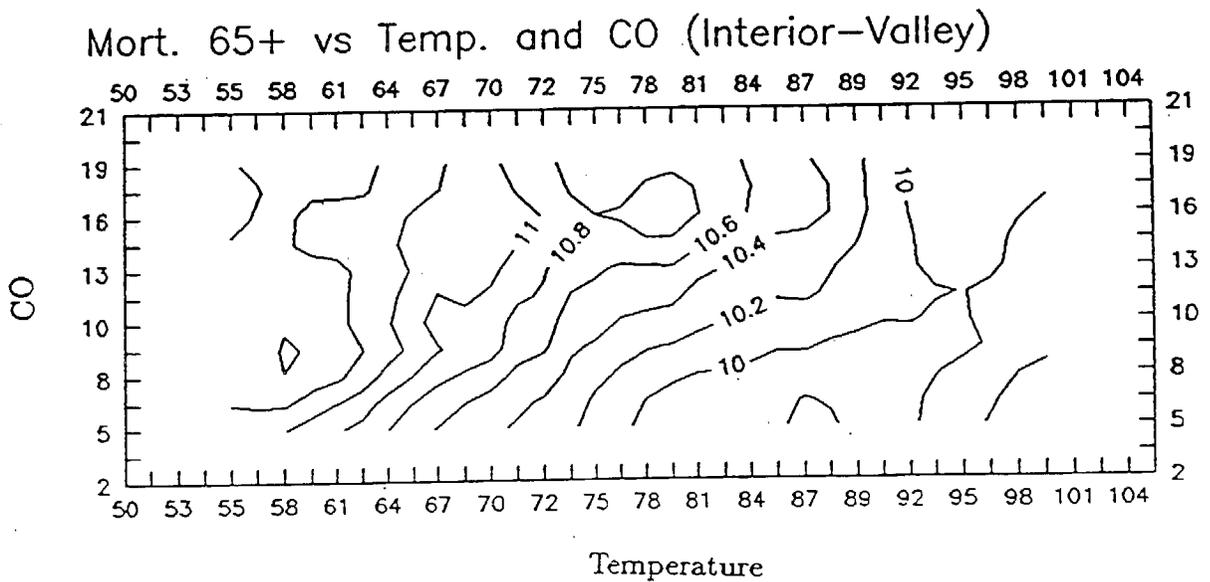
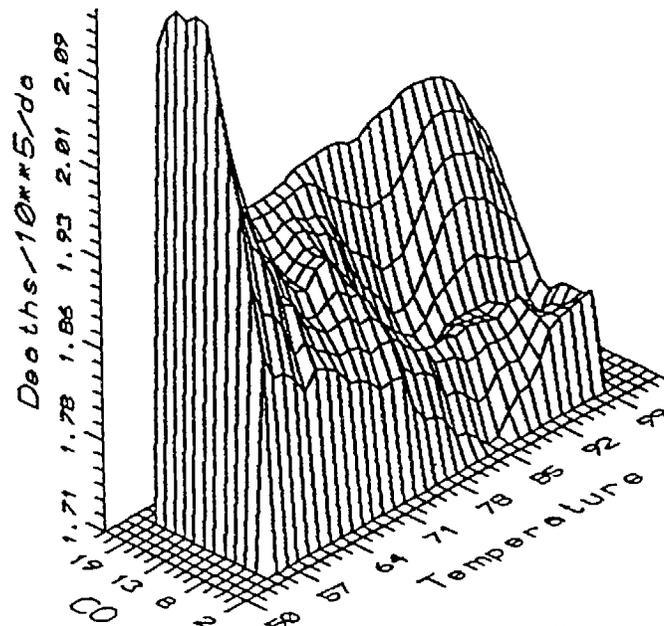


Figure 16a: Mortality Rate (deaths/100,000/day) Profiles and Contours With Temperature and CO for Age 65+ for Interior-Valley Section (1970-1979, 721 5-day intervals).



Mort. 45-64 vs Temp. and CO (Interior-Valley)

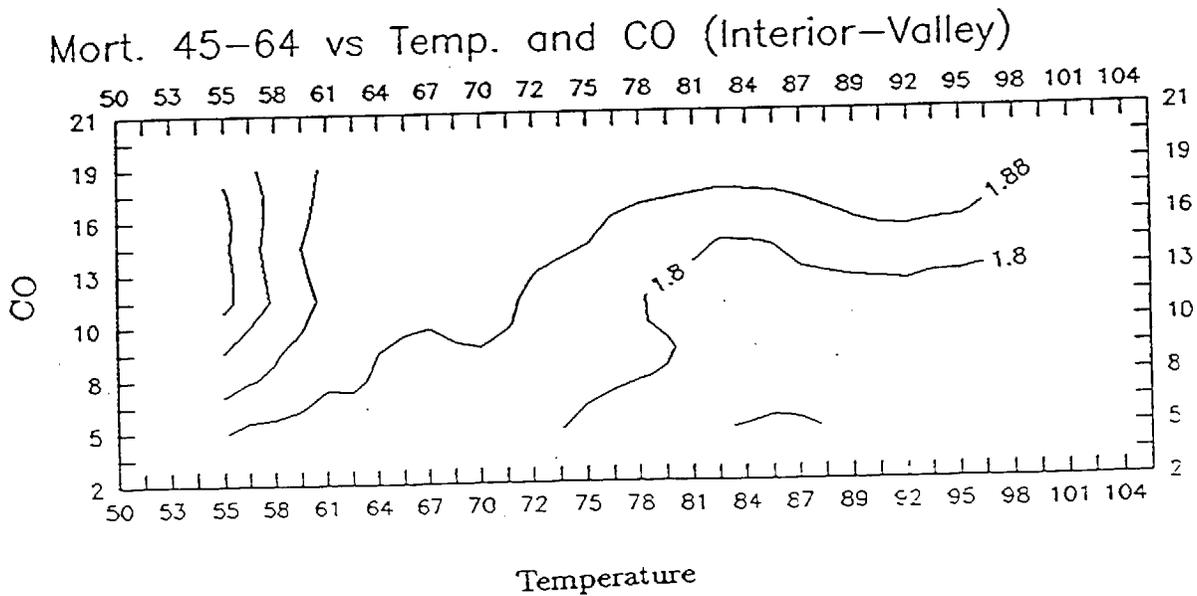
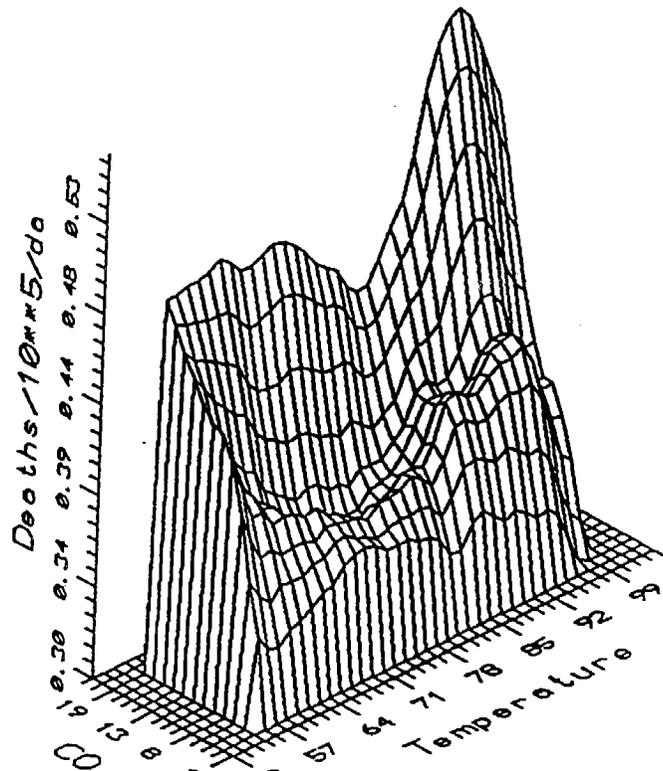


Figure 16b: Mortality Rate (deaths/100,000/day) Profiles and Contours With Temperature and CO for Age 45-64 for Interior-Valley Section (1970-1979, 721 5-intervals).



Mort. 0-10 vs Temp. and CO (Interior-Valley)

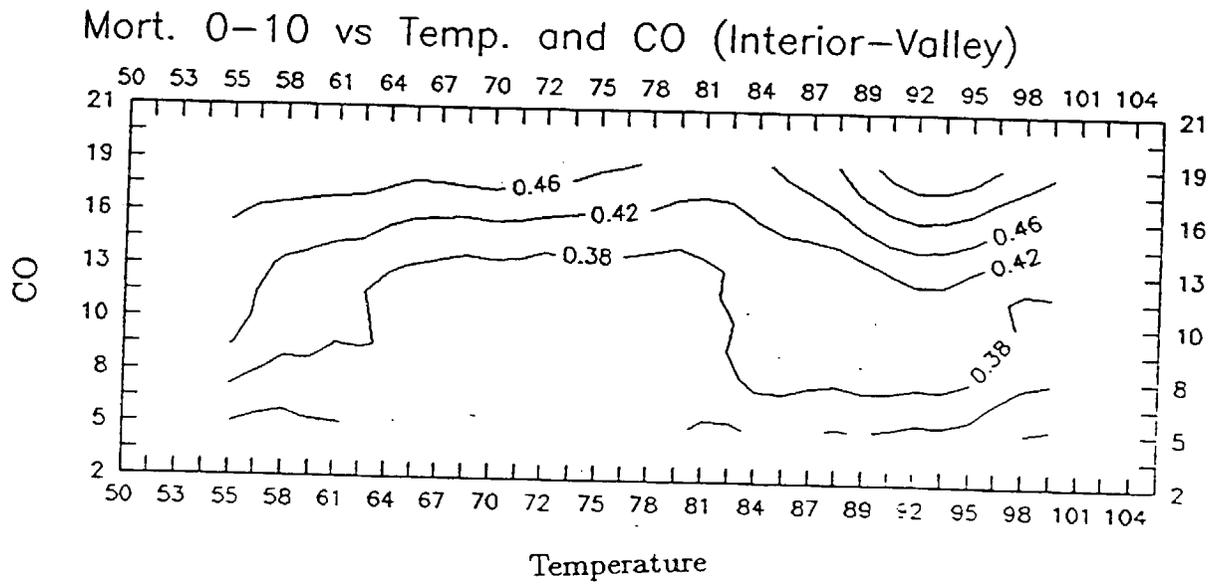
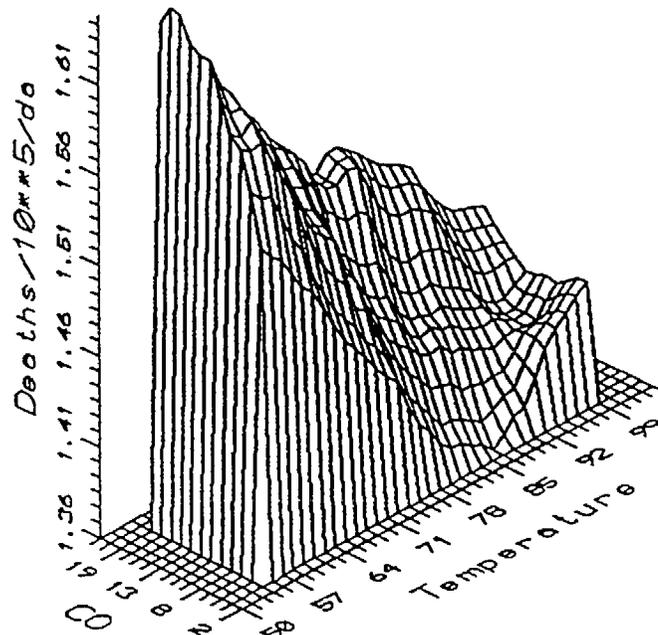


Figure 16c: Mortality Rate (deaths/100,000/day) Profiles and Contours With Temperature and CO for Age 0-9 for Interior-Valley Section (1970-1979, 721 5-day intervals).



Total Mort. vs Temp. and CO (Interior-Valley)

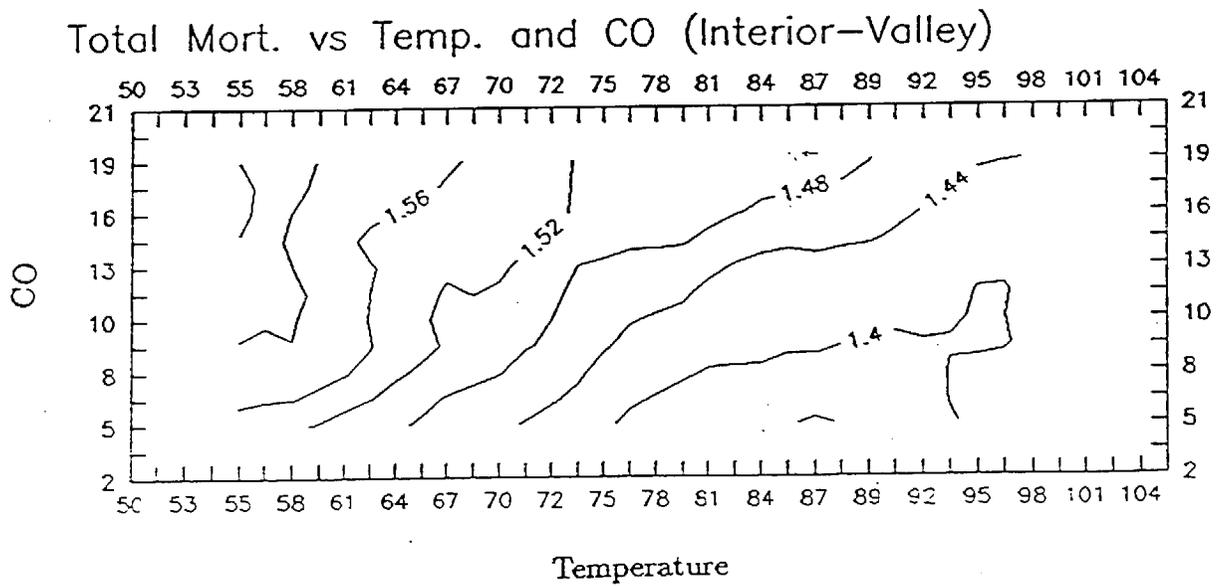
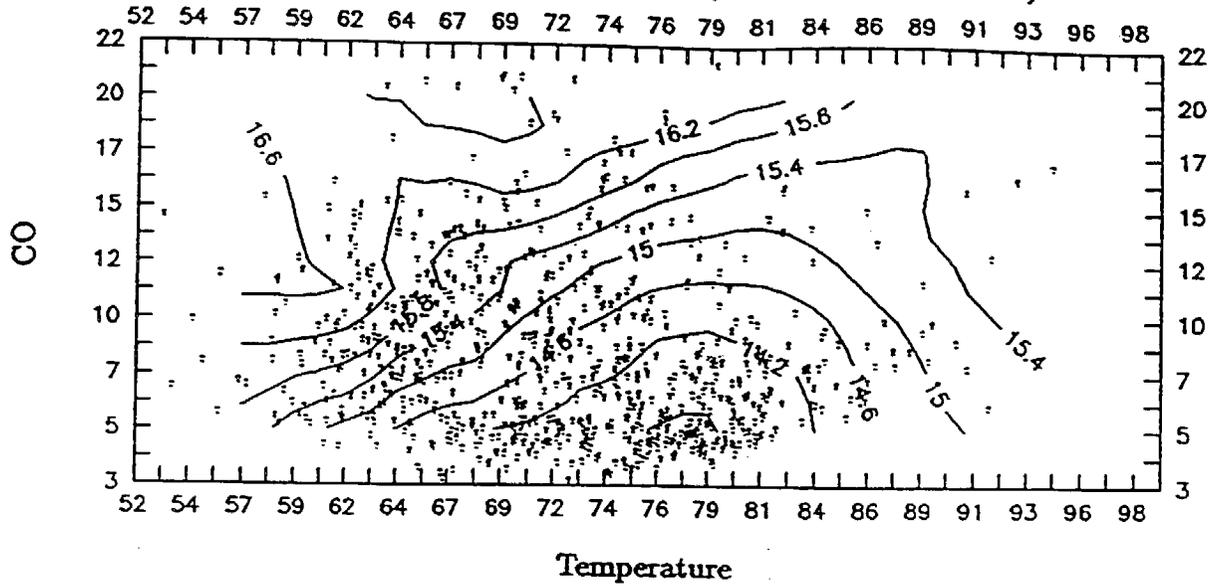


Figure 16d: Mortality Rate (deaths/100,000/day) Profiles and Contours With Temperature and CO (Total) for Interior-Valley Section (1970-1979, 721 5-day intervals).

### Mort. 65+ vs Temp. and CO (Central-Coastal)



### Mort. 65+ vs Temp. and CO (Interior-Valley)

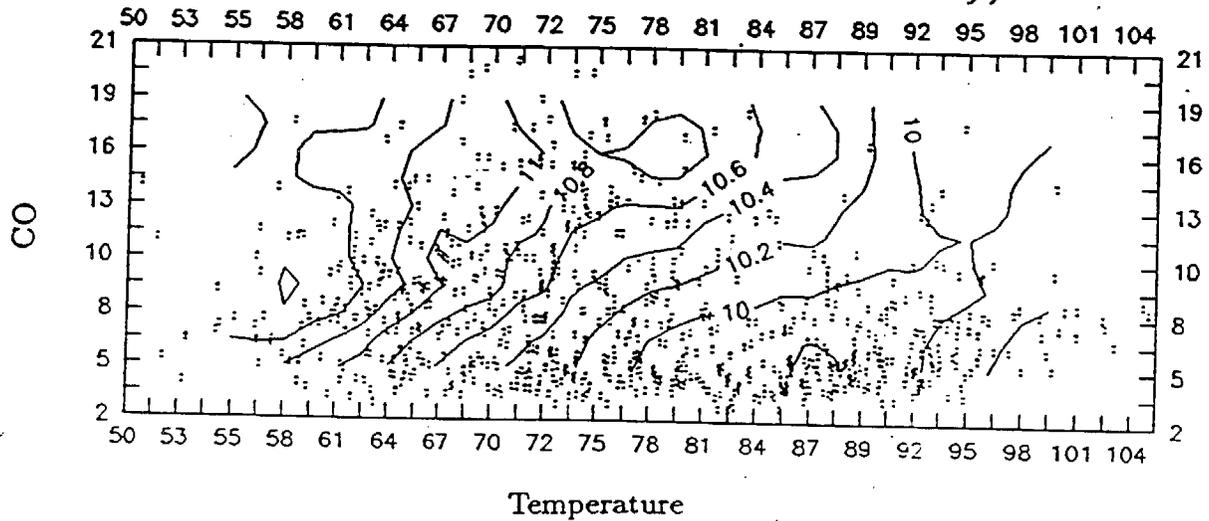


Figure 17: Original Points Superimposed on the 65+ Mortality Profiles for the Central-Coastal and Interior-Valley Populations (Larger dot is data point, smaller dot is label).

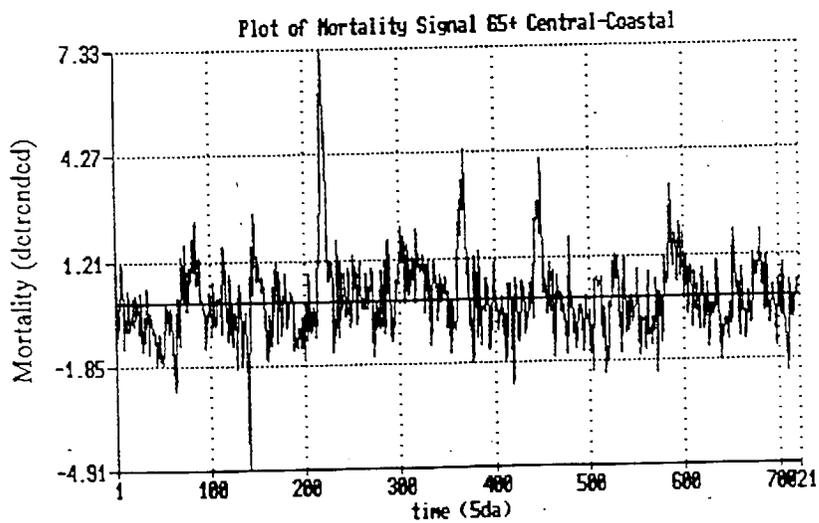
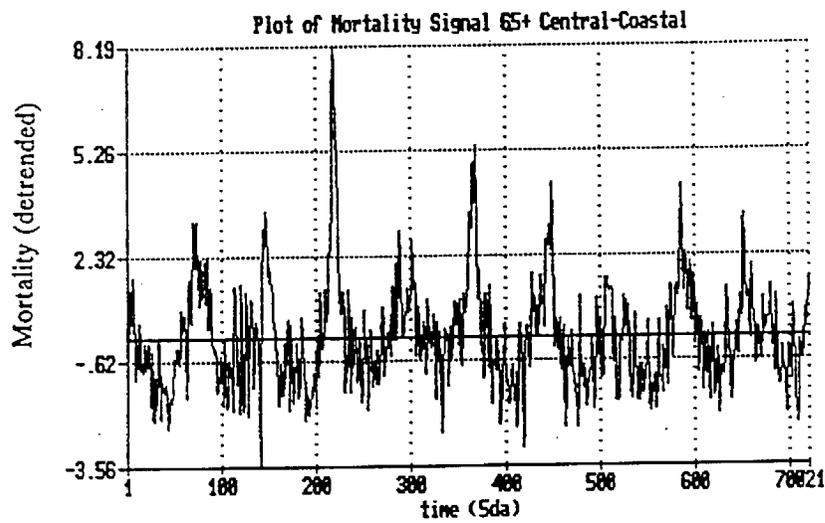
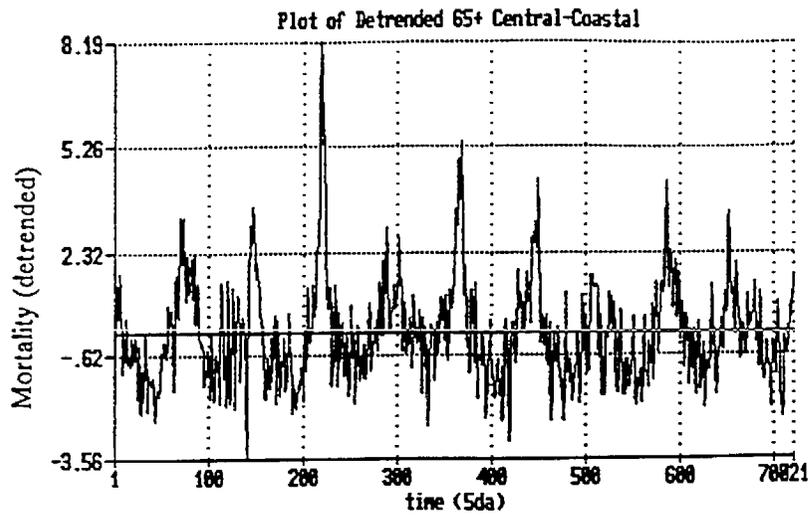


Figure 18: Original Detrended 65+ Central-Coastal Mortality Compared With Estimated Inherent Mortality Signal Estimated by Model I (middle panel) and Model II (bottom panel) Models.

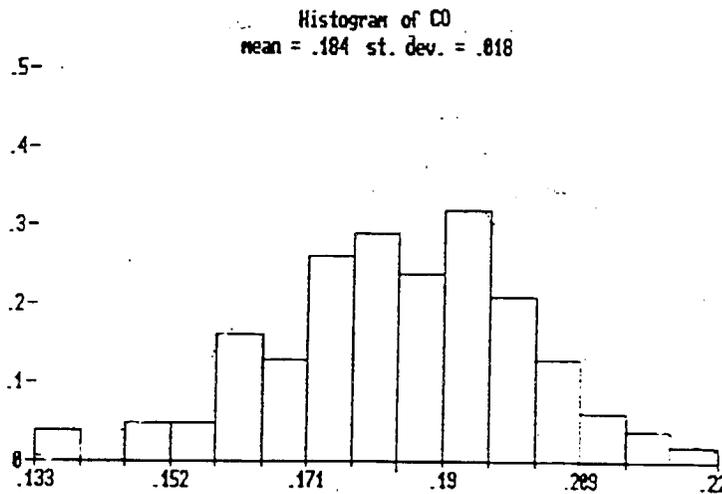
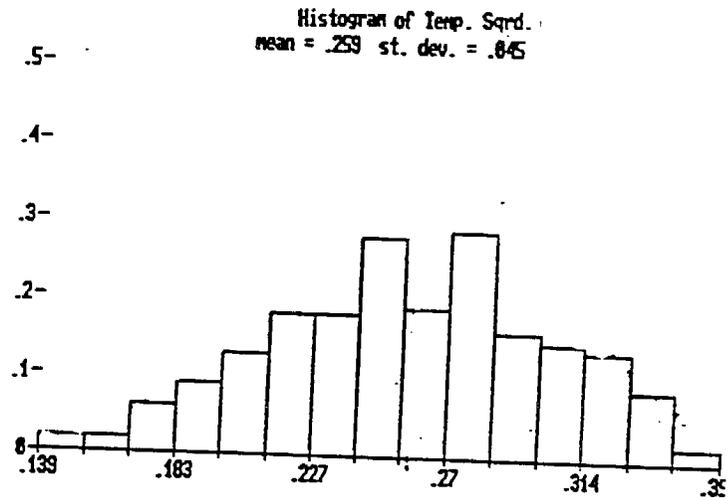
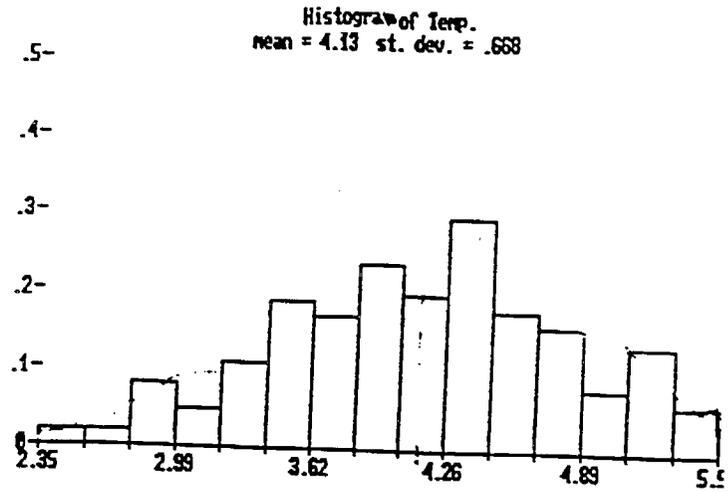
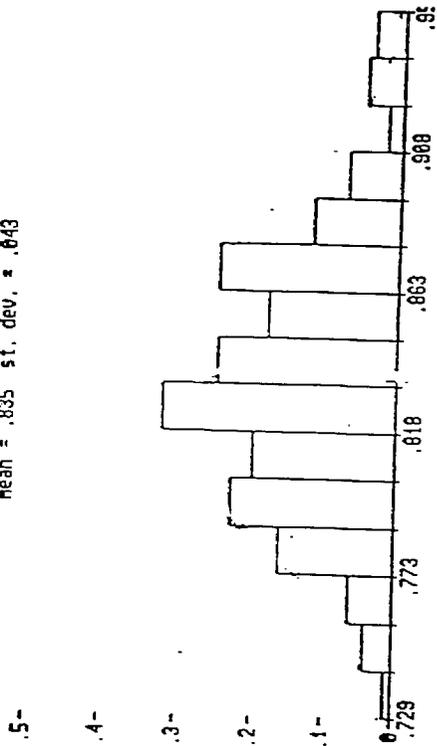
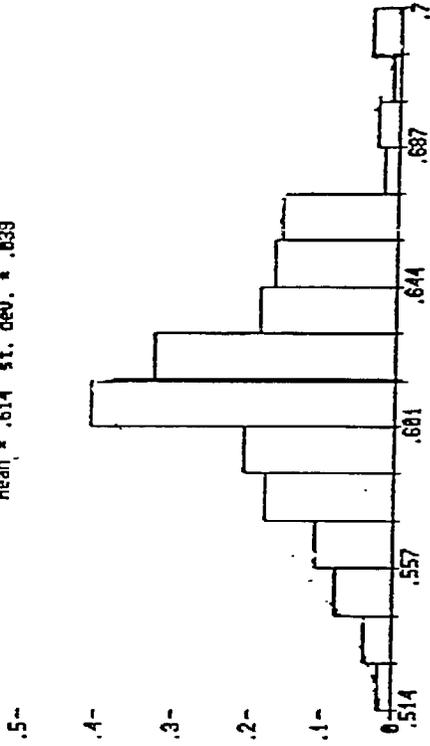


Figure 19: Bootstrap Sampling Distributions of the Coefficients of Temperature ( $\times 10$ ), Squared Temperature ( $\times 100$ ), CO, the Two Estimated Autoregressive Coefficients and the Two Variances.

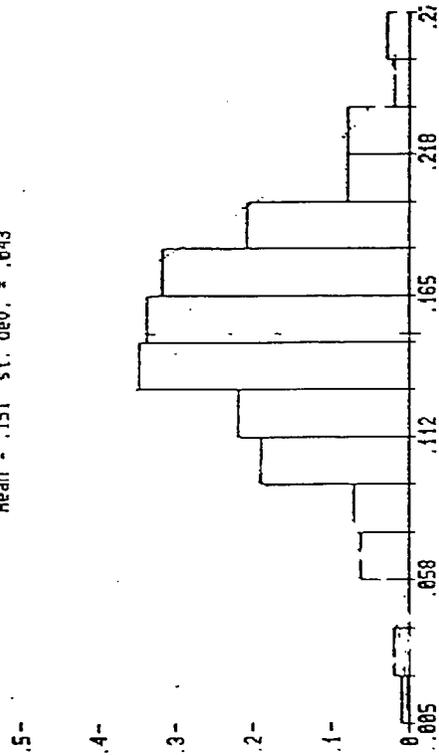
Histogram of AR1  
 mean = .835 st. dev. = .043



Histogram of Model Var.  
 mean = .614 st. dev. = .039



Histogram of AR2  
 mean = .151 st. dev. = .043



Histogram of Obs. Var.  
 mean = .035 st. dev. = .003

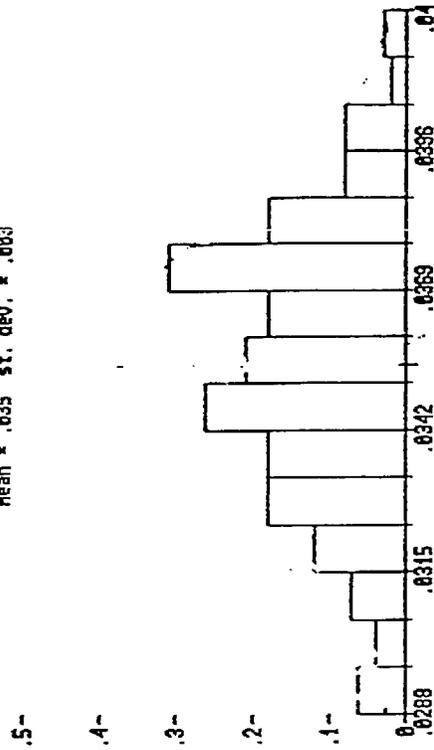


Figure 19: Bootstrap Sampling Distributions of Coefficients of Temperature ( $\times 10$ ), Squared Temperature ( $\times 100$ ), CO, the Two Estimated Autoregressive Coefficients and the Two Variances.

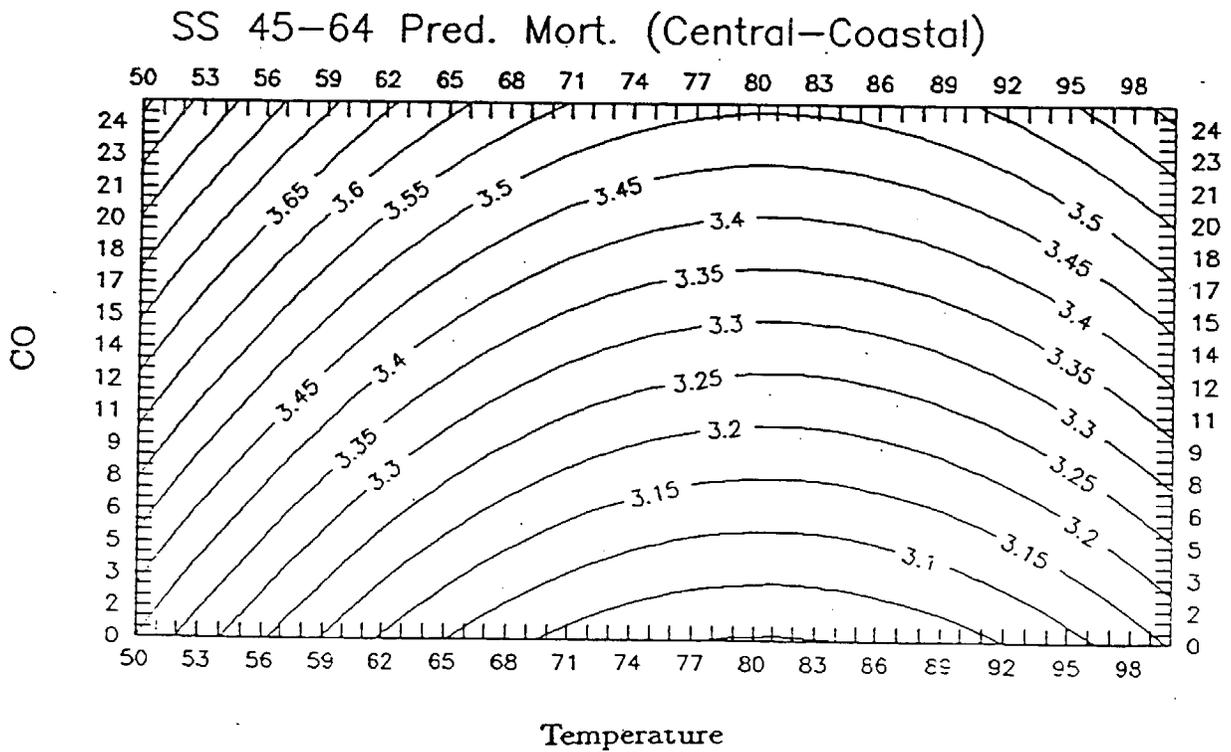
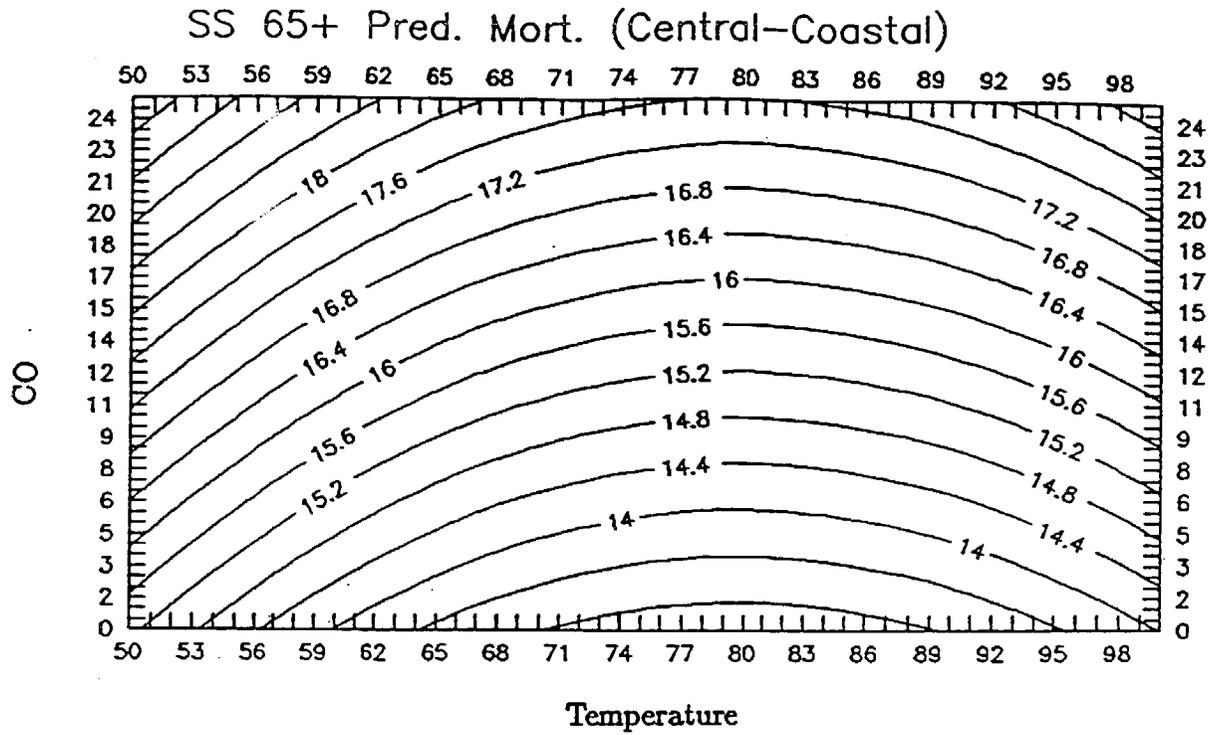


Figure 20a: Total Predicted Mortality (deaths/day) Contours by Age and Section as a Function of Temperature and CO Using Parametric State-Space Model (1970-1979, 721 5-day intervals).

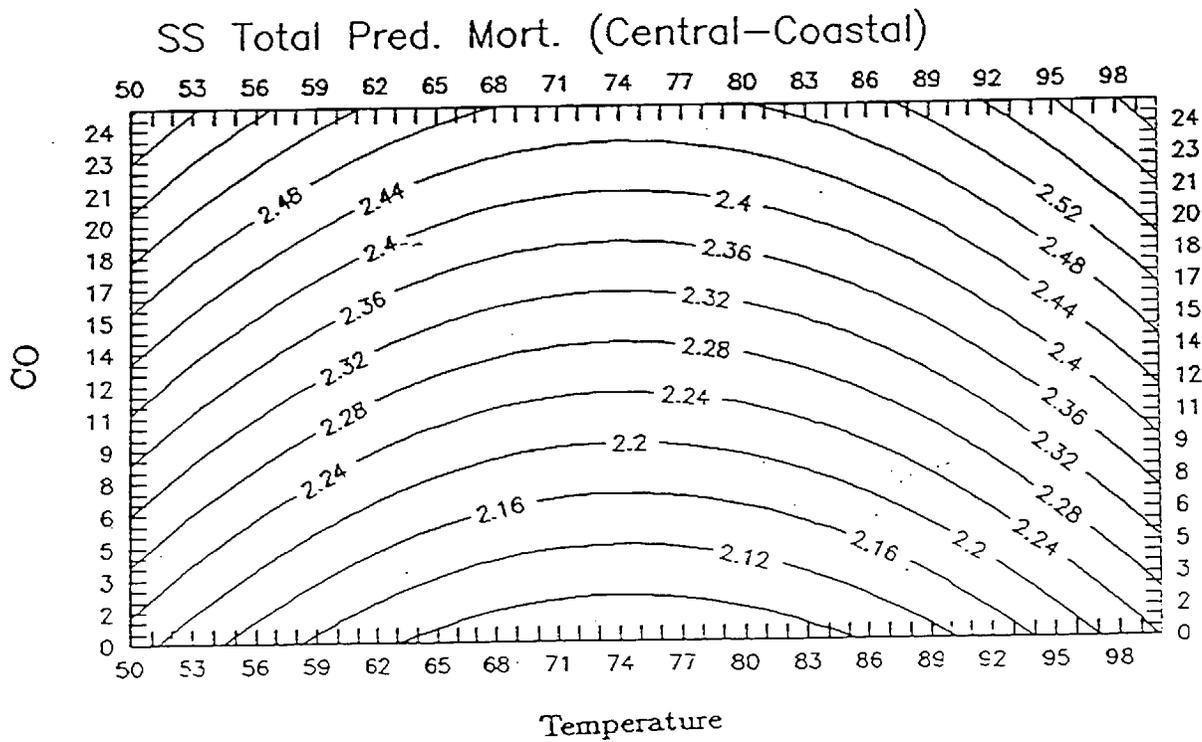
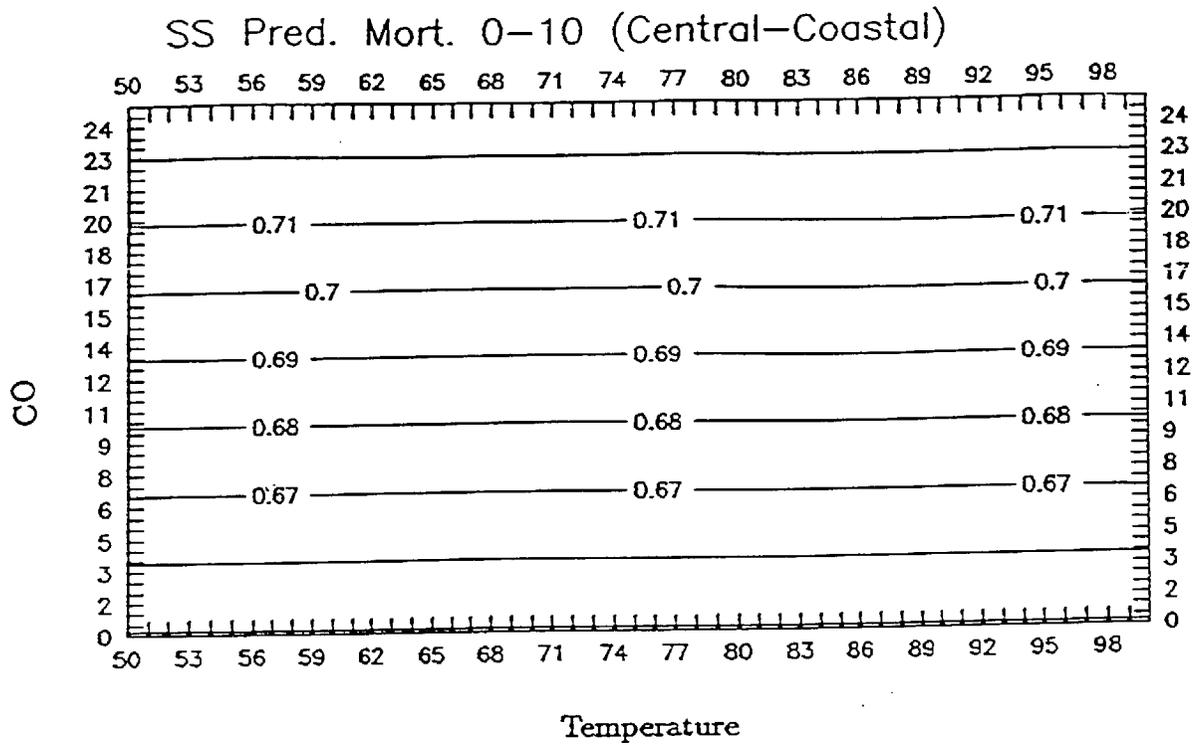


Figure 20b: Total Predicted Mortality (deaths/day) Contours by Age and Section as a Function of Temperature and CO Using Parametric State-Space Model (1970-1979, 721 5-day intervals).

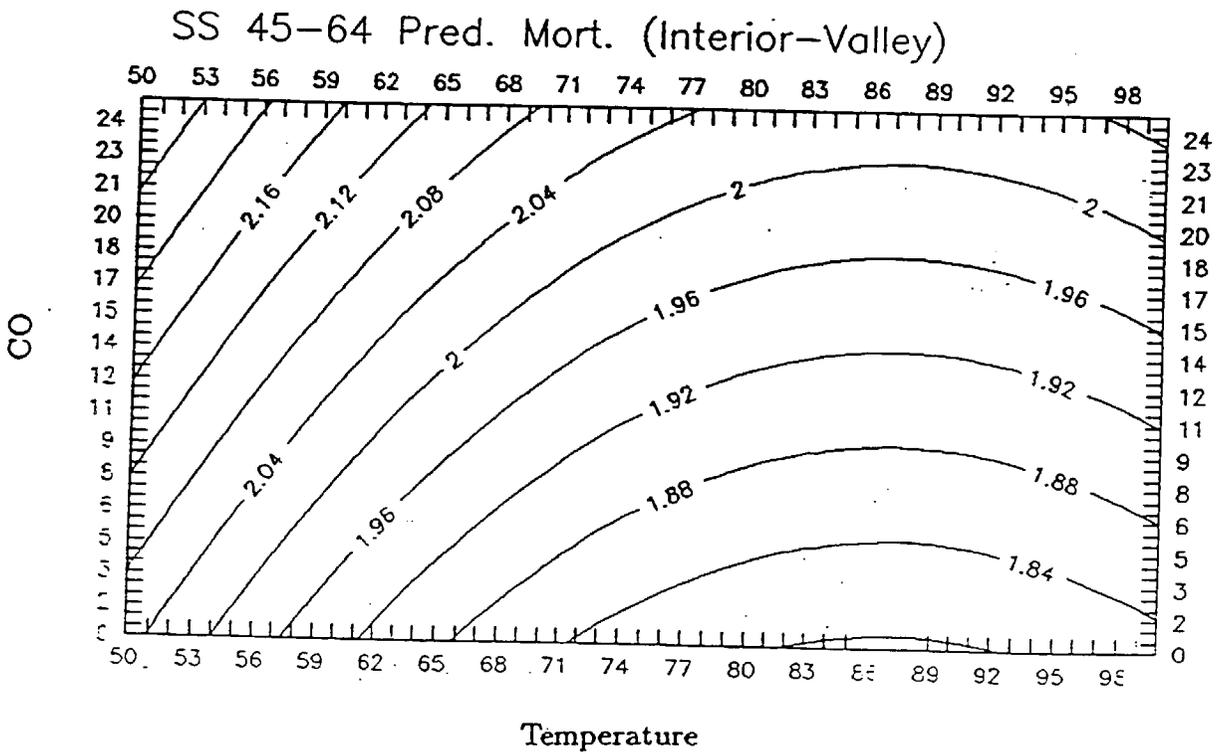
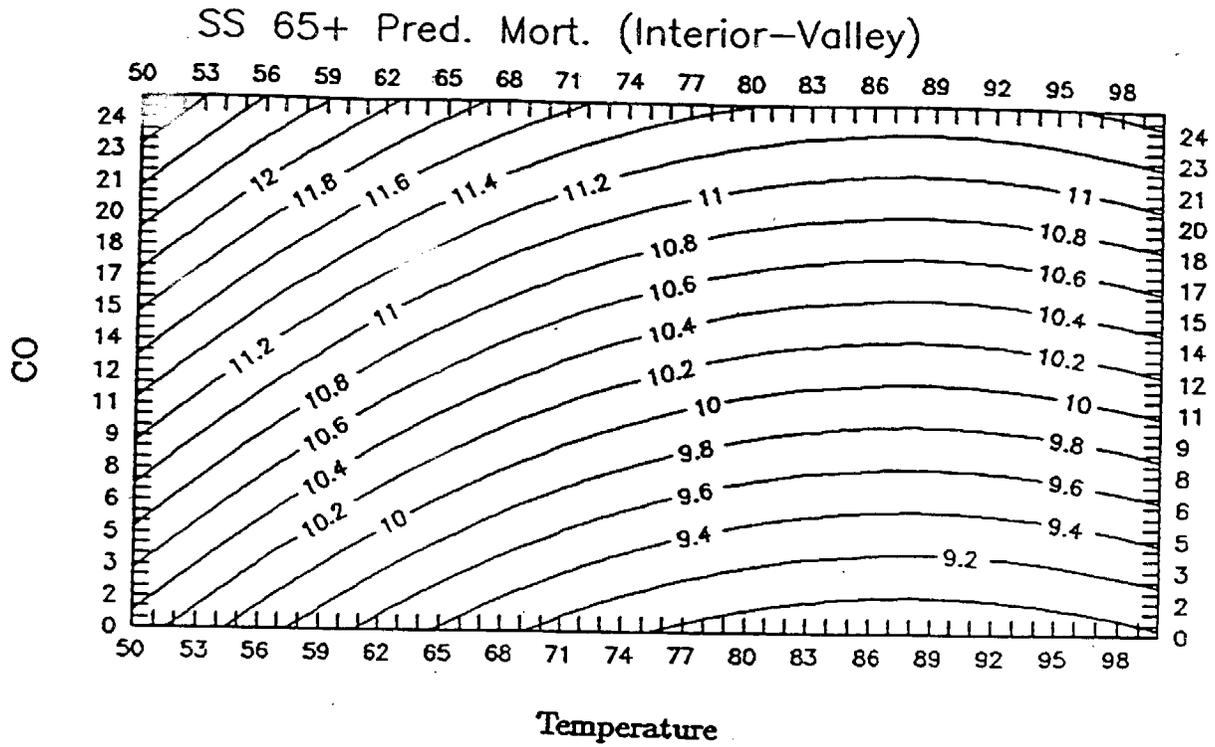
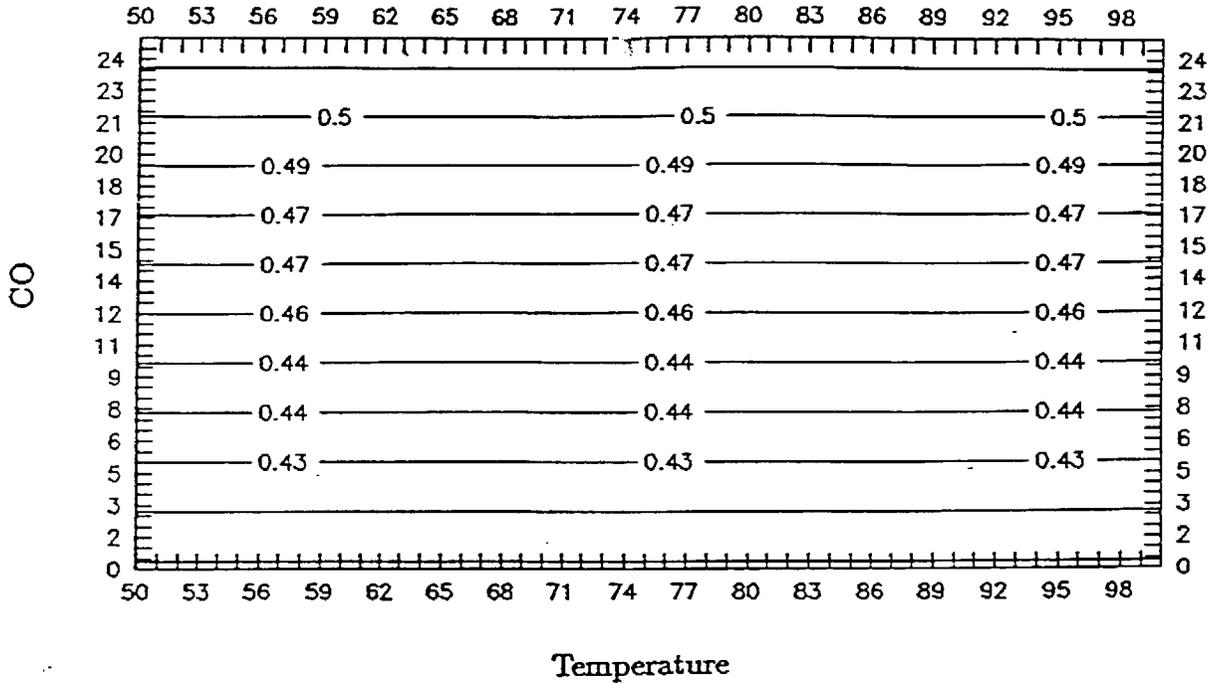


Figure 20c: Total Predicted Mortality (deaths/day) Contours by Age and Section as a Function of Temperature and CO Using Parametric State-Space Model (1970-1979, 721 5-day intervals).

### SS Pred. Mort. 0-10 (Interior-Valley)



### SS Total Pred. Mort. (Interior-Valley)

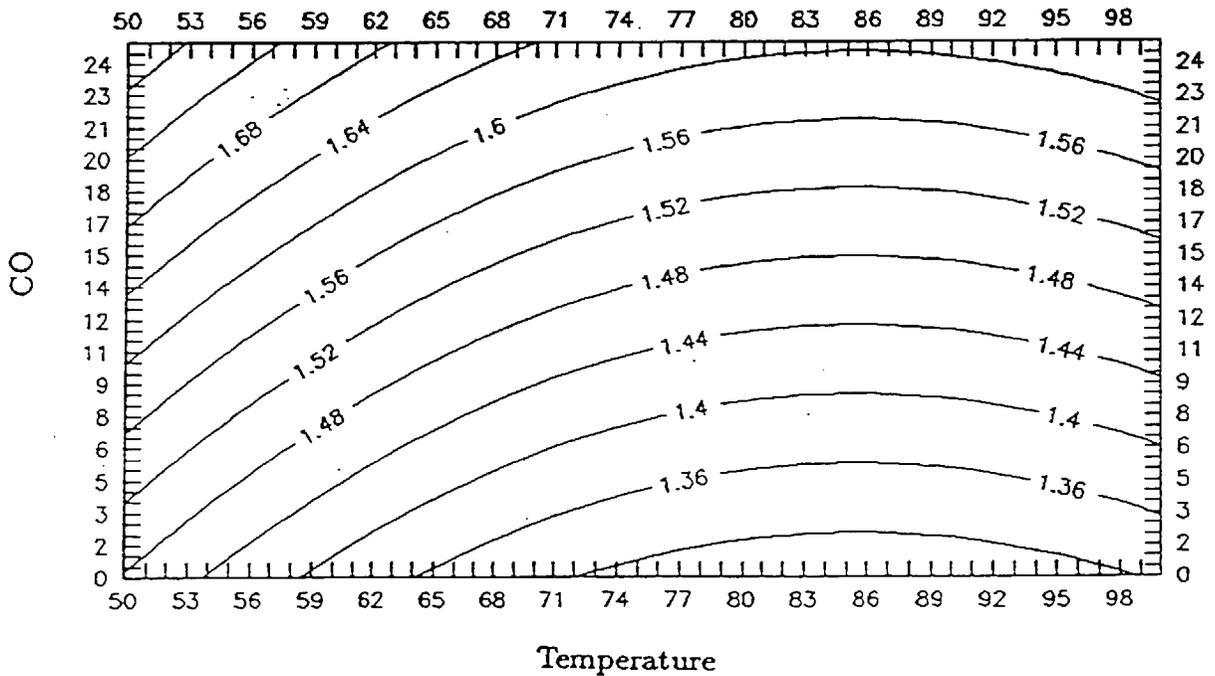


Figure 20d: Total Predicted Mortality (deaths/day) Contours by Age and Section as a Function of Temperature and CO Using Parametric State-Space Model (1970-1971, 721 5-day intervals).

